Optimization of Wireless Networks : Freshness in Communications

Ali Maatouk

université
PARIS-SACLAY

Optimization of Wireless Networks: Freshness in Communications

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 580, Sciences et Technologies de l'Information et de la Communication (STIC)

Spécialité de doctorat: réseaux, information et communications
Unité de recherche: Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France
Référent: Faculté des Sciences d'Orsay

Thèse présentée et soutenue en visioconférence totale, le 16 Novembre 2020, par

Ali MAATOUK

Composition du jury:

Michel Kieffer Professeur, Université Paris-Saclay	Président
Roy Yates Professeur, Rutgers University	Rapporteur & examinateur
Petar Popovski Professeur, Aalborg University	Rapporteur & examinateur
Eytan Modiano Professeur, Massachusetts Institute of Technology	Examinateur
Elif Uysal Professeure, Middle East Technical University	Examinatrice
Mohamad Assaad Professeur, CentraleSupélec	Directeur de thèse
Anthony Ephremides Professeur, University of Maryland	Co-directeur de thèse

Thèse de doctorat

NNT: 2020UPASG038

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Optimisation des Réseaux Sans-Fil: Fraîcheur en Communication

Mots clés: Âge de l'information, Paquets de Mise à Jour, Estimation à Distance, Âge de l'Information Incorrecte, Accès Multiple Avec Écoute de la Porteuse, CSMA, Planification Distribuée, Mise en File d'Attente Prioritaire

Résumé: La prolifération des smartphones, avec la connectivité omniprésente et le bas coût des matériaux, a ouvert la voie à de nouvelles applications qui reposent sur la livraison en temps opportun des paquets d'un bout à l'autre du réseau. De la surveillance des appareils ménagers à la maison au réseaux de véhicules où les informations de vitesse et de position du véhicule sont diffusées, ces applications nécessitent de données frais pour avoir des performances optimales. Pour quantifier cette notion de fraîcheur de données, le concept de l'Âge de l'Information (AdI) est né, et la recherche s'est fortement concentrée sur son analyse et son optimisation dans divers contextes de réseau. Cette thèse explore l'AdI dans de nombreux environnements, met en lumière ses points faibles et leur apporte ainsi des solutions dans plusieurs applications de surveillance en temps réel.

Dans la première partie de la thèse, nous nous concentrons sur l'optimisation des métriques basées sur l'âge dans les systèmes de communication fondamentaux. Plus précisément, dans le troisième chapitre, nous examinons les métriques basées sur l'âge dans les environnements multi-classes qui sont abondants dans les applications en temps réel. Un exemple simple est celui des réseaux de véhicules où les données relatives à la sécurité sont considérées comme plus sensibles. Par conséquent, elles ont une priorité plus élevée par rapport aux autres données du système. Nous dérivons une expression de l'âge moyen de chaque flux et nous fournissons des résultats à propos de l'interaction entre les multiples classes. Cela ouvre la voie à la deuxième partie du chapitre, où nous introduisons un nouveau cadre d'optimisation basé sur l'AdI dans les systèmes multi-classes. Nous y caractérisons les gains en termes de fraîcheur de l'information lorsque notre cadre est adopté par rapport à des approches de pointe. Le quatrième chapitre traite un environnement distribué, où les appareils accèdent au canal en utilisant la méthode d'accès multiple avec écoute de la porteuse (CSMA). CSMA est considéré comme l'un des méthodes d'accès canal distribués les plus connus et les plus répandus (par exemple, CSMA est le principal moyen d'accès en Wi-Fi). Dans ce cas, nous caractérisons, grâce à des analyses théoriques rigoureuses, le point de fonctionnement optimal qui minimise l'âge moyen du réseau.

Dans la deuxième partie de la thèse, nous mettons en lumière les lacunes de l'âge de l'information et des métriques d'erreur standard dans de nombreuses applications en temps réel. Par conséquent, nous introduisons une nouvelle métrique de performance, que nous appelons l'Âge de l'Information Incorrecte (AdII). L'AdII traite ces lacunes en étendant la notion de données frais et en saisissant l'effet de détérioration que les informations incorrectes peuvent avoir avec le temps sur le système. Dans les scénarios à la fois sans et avec contraintes de ressources, nous dérivons des politiques d'échantillonnage optimales qui minimisent l'AdII. Nous soulignons également leurs avantages par rapport aux politiques optimales pour l'âge et pour les métriques d'erreur standard dans diverses applications. Nos résultats et analyses fournissent des informations clés sur la métrique d'âge et ouvrent la voie à de nouvelles orientations de recherche pour les applications de surveillance en temps réel.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Title: Optimization of Wireless Networks: Freshness in Communications

Keywords: Age of Information, Status Updates, Remote Estimation, Age of Incorrect Information, Carrier Sense Multiple Access, CSMA, Distributed Scheduling, Priority Queuing

Abstract: The proliferation of smartphones, along with the ubiquitous connectivity and cheap hardware cost, has paved the way for new applications that rely on the timely delivery of packets from one end of the network to another. From monitoring home appliances back at the house to vehicular networks where the vehicle's velocity and position information are disseminated, these applications require fresh data to have optimal performance. To quantify this notion of freshness, the concept of the Age of Information (AoI) was born, and research attention has been put heavily on its analysis and optimization in various network settings. This thesis explores the AoI in numerous system environments, sheds light on its shortcomings, and accordingly provides solutions to them in several real-time monitoring applications.

In the first part of the thesis, we focus on optimizing age-based metrics in fundamental communication systems. Specifically, in the third chapter, we examine age-based metrics in multi-class environments that are abundant in real-time applications. A simple example is vehicular networks where safety-related data are considered more sensitive. Consequently, they have a higher priority than the other data in the system. We derive a closed-form expression of each stream's average age and provide substantial insights into the interaction between the multiple classes. This paves the way for the second part of the chapter, where we introduce a new AoI-based optimization framework in multi-class systems. Therein, we characterize the gains in terms of information freshness when our framework is adopted compared to state-of-the-art approaches. The fourth chapter deals with a distributed scheduling environment, where devices contend for the channel using the well-known carrier sense multiple access scheme (CSMA). CSMA is considered one of the most renowned and widely spread distributed scheduling schemes (e.g., CSMA is the primary medium access in Wi-Fi). We characterize, through rigorous theoretical analyses, the operating point that minimizes the average AoI.

In the second part of the thesis, we shed light on the shortcomings of the age of information and standard error metrics in many real-time applications. Toward that end, we introduce a new performance metric, which we refer to as the Age of Incorrect Information (AoII). AoII deals with these shortcomings as it extends the notion of fresh updates and captures the deteriorating effect wrong information can have with time on the system. In both unconstrained and resource-constraint scenarios, we derive optimal sampling policies that minimize the AoII. We also highlight their advantages compared to both the age-optimal and error-optimal policies in a variety of real-life applications. Our results and analyses provide key insights into the age metric and lead the way to novel research directions for real-time monitoring applications.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Dedicated to my family for their unlimited support

Acknowledgments

I want to take this opportunity to express my heartfelt gratitude to all those who helped me to make my thesis work a success. Foremost, my thesis advisor **Mohamad Assaad**, whose constant support was the critical element to this success. Ever since I was a master's student, he provided me with guidance and countless advice on how to choose my thesis topic. The door to his office was always open during my thesis. He offered me generously his time, to hear about the ideas I have in mind or the possible solutions to the problems I am facing in my research. I am lucky to have a supervisor who cared so much about my work, and who always thought outside the box and encouraged me to keep my work pace as high as possible. I am also grateful for all our non-technical discussions, especially about the socio-economic issues in the world, which helped me better understand these problems.

My sincere appreciation and gratitude go to **Anthony Ephremides**, my co-advisor. Being such a big name in the community, I remember being nervous when I met him for the first time. But, thanks to his friendly personality and welcoming attitude, the thesis' work quickly kick-started as we shared our ideas on the topic at hand. I also credit this to his unrivaled enthusiasm; it was so contagious as I would always walk out of every meeting with him, fueled with motivation and desire to work harder than ever. I also appreciate his efforts to introduce me to as many of the community members as possible at each international conference. I will always remember our conversations about Europe's history and on which region produces the best wine.

I would also like to express my appreciation to the members of the jury committee: **Michel Kieffer**, **Roy Yates**, **Petar Popovski**, **Eytan Modiano**, and **Elif Uysal** for the time spent reading this work and for all their valuable feedback on it.

I would like to thank **Yin Sun**, a great collaborator I was lucky enough to work with. He always sought perfection in mathematical writing, provided me with videos and articles on how to master the art of mathematical writing, which was immensely useful to my thesis work. He also never hesitated to provide me with insights and his perspective on my future career choices.

Thanks also go out to my office colleague **Saad Kriouille** with whom I spent numerous hours trying to solve the issues faced in our collaborative works. The same goes for my colleagues **Salah Eddine Hajri** and **Juwendo Denis**, not only for their technical contributions but also for being instrumental in maintaining a relaxed atmosphere around the department.

I would also like to acknowledge the professors whose working along their side was essential to my self-development. **Hikmet Sari**, whose witnessing his work ethics was important to develop my approach to research. **Abdellatif Samhat** for

helping me in my journey from Lebanon to France. **Meroune Debbah** for his motivational conversations and providing me with guidance on the current trends of research.

I also want to express my gratitude to my cousin **Ramzi Cheaito**, and all my friends in Paris who made it become a second home for me. **Hassan Abou Ibrahim**, **Océane Cayuela**, **Mathis Reguilon**, **Eloise Letourneur**, **Marianne Auguste**, **Essra Abdallah**, and **Florent Poubanne**, you have been an invaluable part of my thesis journey. Living in a new country is not straightforward, but the time spent together surely made things easier.

Lastly, it is always the most difficult to express appreciation to the people closest to you. **Janaki Goudar**, who always believed in me, inspired me and stuck with me in my hardest times. Thank you for always listening to my rambling about the mathematical issues I am facing in my work. I have, still, and will forever cherish the countless work dates that made writing my research papers easier to handle. There are no words to express my gratitude and thanks to my beloved parents and family members in Lebanon for always standing by me. Thank you for your unconditional love, having my back, and handling the burden of me living in another country. Knowing that I could always turn to you for advice, encouragement, and support made this Ph.D. journey possible. I am forever in debt to you.

Contents

List of Figures

List of Tables

Acronyms

ACK Acknowledgement. 8, 24

AdI Âge de l'Information. 5

AdII Âge de l'Information Incorrecte. 13

AdS Âge de Synchronisation. 11

AoI Age of Information. 21

AoII Age of Incorrect Information. 28

AoS Age of Synchronization. 27

CMDP Constrained Markov Decision Process. 90

CSMA Carrier Sense Multiple Access. 10, 26

CW Contention Window. 65

DTMC Discrete Time Markov Chain. 102

FCFS First-Come-First-Served. 5, 21

IdO Internet des Objets. 2

iid Independent and Identically Distributed. 95

IoT Internet of Things. 18

KKT Karush-Kuhn-Tucker. 144

LGFS Last-Generated-First-Served. 8, 23

LHS Left Hand Side. 161

LTI Linear Time-Invariant. 11, 26

MAC Medium-Access-Control. 28

MAdI Maximum Âge de l'Information. 6

MAF-LGFS Maximum Age First, Last-Generated, First-Served. 51

MDP Markov Decision Process. 9, 25

MMSE Minimum Mean Squared Error. 11, 26

MSE Mean Squared Error. 11, 26

NACK Negative Acknowledgement. 95

NCS Network Controlled Systems. 11, 26

PAoI Peak Age of Information. 22

PP-MAF-LGFS Preemptive-Priority, Maximum-Age-First,Last-Generated First-Served. 12, 28

RHS Right Hand Side. 153

RR Round Robin. 10, 26

RSUs Road Side Units. 18

SCA Sequential Convex Approximation. 58

SHS Stochastic Hybrid System. 34

SINR Signal-to-Interference-plus-Noise Ratio. 10, 25

UAV Unmanned Aerial Vehicles. 22

V2X Vehicle-To-Everything. 52

VIA Value Iteration Algorithm. 153

1 Introduction (French)

1.1 Contexte et Motivation

Naturellement, les êtres humains sont une espèce sociale; nous partageons des neurones miroirs qui nous permettent de faire correspondre les émotions de chacun immédiatement et inconsciemment. Nous reflétons même l'activité cérébrale de l'autre lorsque nous sommes engagés dans la narration. Nous créons des liens, exprimons des émotions et formons des sociétés. Dans son célèbre livre «Politique», Aristote décrit les humains comme suit:

L'homme est naturellement un animal social.

Ces observations fondamentales montrent que personne ne peut briser les chaînes de la dépendance mutuelle. Les humains doivent satisfaire leurs besoins sociaux naturels et avoir leur sentiment d'appartenance. Que ce soit sa maison, sa communauté ou son état; un environnement social est indispensable à la stabilité d'un individu. Les sociologues considèrent que cela commence dans la relation entre l'embryon et la mère et se poursuit jusqu'à la mort.

Au cœur de nos interactions sociales se trouve le concept fondamental de *communication*. La racine du mot communication est le mot latin *communicare*, qui signifie partager ou rendre commun. À cette fin, la communication est définie comme un processus entre au moins deux entités où a lieu un transfert d'informations d'une entité à l'autre. La communication est l'épine dorsale de toutes les relations humaines. Cela nous aide à diffuser les connaissances et nous permet d'exprimer nos idées et nos sentiments et de comprendre les émotions et les pensées des autres. En conséquence, l'affection est développée et des relations positives et négatives peuvent être fondées. Par conséquent, nous ne pouvons pas sous-estimer l'importance de la communication. Dans sa forme la plus simple, et comme le montre la Fig. 1.1, tout processus de communication comporte trois éléments essentiels:

1. **Source**: La source est l'entité qui crée et envoie le message. Il ou elle commence par décider d'abord quel message à communiquer. Le message est ensuite codé en déterminant la combinaison parfaite d'expressions verbales (par exemple, des mots) et non verbales (par exemple, langage corporel, ton de la

voix) qui transmettent la signification voulue. Enfin, le message se propage à travers un support afin d'atteindre le destinataire.

2. **Canal**: Le canal est le moyen par lequel le message codé circule entre la source et le récepteur. Par exemple, dans des situations parlées comme les conversations en face à face, le médium est l'air dans lequel les ondes sonores mécaniques se propagent. Le canal est généralement soumis à plusieurs perturbations qui peuvent gêner le processus de communication. Par exemple, dans la même conversation en face à face, le bruit du vent ou la voix des autres orateurs peuvent affecter le processus de communication en cours.

3. **Destinataire**: Le destinataire est l'entité à laquelle le message est destiné. Après avoir traversé le canal, le message est décodé et analysé par le récepteur. La mesure dans laquelle le destinataire comprend le message dépend fortement de nombreux facteurs. Parmi ces facteurs, nous citons le bruit et l'interférence dans le canal, et le temps de réception du message.

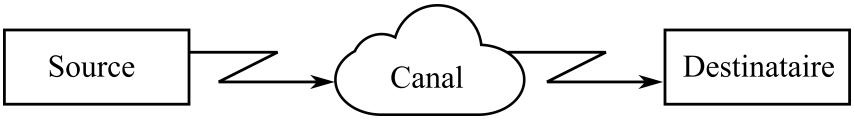


Figure 1.1: Processus de communication de base.

Poussés par leurs besoins sociaux, les humains ont continué à être les pionniers de nouvelles formes de communication au-delà des conversations en face à face standard. De l'utilisation de l'écriture hiéroglyphique dans l'Égypte ancienne à l'invention de l'alphabet par les Phéniciens, l'innovation humaine à cet égard n'a jamais cessé. Au cours des deux derniers siècles, les progrès de la technologie ont créé un environnement fertile à partir duquel a émergé un monde connecté où la communication de presque n'importe quel point de la terre (et au-delà) à l'autre est à la fois faisable et rapide. De la diffusion par satellite à la prolifération des téléphones intelligents, le monde n'a jamais été aussi connecté qu'aujourd'hui. Toutes ces avancées récentes dans les systèmes de communication ont radicalement changé la façon dont les humains communiquent et accèdent/échangent les informations. Ce changement radical a atteint un point où la communication a dépassé l'exclusivité de la communication interhumaine. Plus précisément, grâce à la connectivité omniprésente et au bas coût des matériaux, l'Internet des Objets (IdO) est né.

Au sens le plus large, un système IdO se compose de dispositifs informatiques interdépendants, de machines mécaniques et numériques capables de transférer des données sur un réseau sans nécessiter d'interaction d'homme à homme ou d'homme à ordinateur [1]. Dans leur rapport de 2013, McKinsey note une croissance de 300 % des appareils connectés IdO au cours des cinq dernières années et évalue le potentiel impact économique de l'IdO entre 2,7 billions de dollars et 6,2 billions de dollars par an en 2025 [2]. Cela montre que les systèmes IdO prennent de l'ampleur dans l'industrie et dans la recherche communautés et son importance dans notre vie

quotidienne future. Nous rapportons ci-dessous divers exemples d'applications de systèmes IdO.

- **Réseaux de Véhicules**: Ces réseaux sont formés par des véhicules en mouvement, des unités côté route, et des piétons qui transportent des appareils de communication [3]. Les messages générés et envoyés par ces entités à travers le réseau contiennent des données telles que la vitesse et la position des véhicules, les feux de signalisation, les mises à jour des accidents, etc. [4].

- **Maisons Intelligentes**: Dans les applications pour maisons intelligentes, une personne peut gérer, surveiller et contrôler à distance divers services du bâtiment tels que la climatisation, les systèmes de divertissement, l'éclairage et les appareils ménagers [5].

- **Surveillance de l'Environnement**: Cette branche d'applications couvre les processus et les activités nécessaires pour identifier et surveiller la qualité de l'environnement (par exemple, la qualité de l'air, de l'eau et du sol). Grâce aux progrès récents du matériel et des technologies sans-fil, cela peut être réalisé en relayant les données de capteurs répartis dans la zone environnementale d'intérêt vers un moniteur à distance pour le traitement [6].

Au fond, les applications ci-dessus et de nombreuses autres applications IdO appartiennent à la même catégorie de services IdO appelée "**Surveillance en temps réel**". Dans ces systèmes, une entité souhaite connaître l'état d'un ou de plusieurs processus observés par une source distante. En conséquence, la source envoie des paquets de *mise à jour d'état* au moniteur pour fournir des informations sur le/les processus d'intérêt. Nous distinguons dans ce cas deux scénarios:

- **Générer à Volonté**: Dans ce scénario, la source distante peut générer des mises à jour d'état à n'importe quelle instance de temps souhaitée [7]. Un exemple est lorsque la source est un capteur de température visant à reproduire le processus de température chez le moniteur.

- **Arrivées Stochastiques**: Dans ce scénario, les mises à jour d'état arrivent aléatoirement à la source selon une distribution de probabilité. C'est le scénario dans lequel la source distante n'a aucun contrôle sur les arrivées des paquets.

Dans les deux scénarios, une fois que le paquet de mise à jour d'état arrive au file d'attente, il est préparé pour la transmission sur le canal. Ensuite, la mise à jour est envoyée via le réseau, qui peut être constitué de plusieurs liaisons filaires/sans-fil. Une fois la mise à jour de l'état est fournie au moniteur, celui-ci peut continuer à exécuter ses tâches. Ce système de mise à jour d'état est rapporté dans la Fig. 1.2. Nous notons que le système actuel de mise à jour d'état n'est rien d'autre qu'un cas particulier du modèle de communication de base représenté dans la Fig. 1.1. Plus précisément, dans le système de mise à jour d'état, la source/récepteur peuvent être

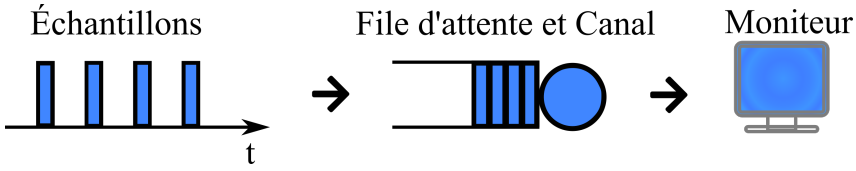


Figure 1.2: Système de mise à jour d'état.

soit des machines, soit des humains, et le canal est une combinaison arbitraire de liaisons filaires et sans-fil.

L'objectif principal de ces applications est de tenir le moniteur informé et à jour sur le processus observé par la source. En d'autres termes, les informations ont la plus grande valeur lorsqu'elles sont fraîches. Plus un paquet est ancien, plus la valeur qu'il apporte au moniteur est faible. Pour mieux comprendre, prenons l'exemple d'un réseau de véhicules où la position des véhicules est diffusée à une entité centrale qui gère les feux de signalisation. Pour réduire les embouteillages et fluidifier la circulation, l'entité centrale contrôle les feux de signalisation en fonction des informations dont elle dispose sur les différents véhicules. De toute évidence, l'efficacité des décisions du contrôleur dépendra fortement de la mise à jour des positions des véhicules. Par exemple, supposons que le contrôleur dispose d'informations sur la position des véhicules il y a 15 minutes. L'algorithme utilisé pour optimiser le trafic ne conduira pas aux mêmes résultats par rapport au cas où le contrôleur dispose d'information sur la position des véhicules il y a seulement 5 secondes. L'algorithme aggravera probablement le trafic en raison de l'utilisation de données obsolètes. Cette observation met en perspective à quel point ces applications dépendent de nouvelles données. Dans cet esprit, nous notons que de la génération à la réception, le paquet de mise à jour d'état peut être sujet à une congestion de la file d'attente, des erreurs de canal et d'autres retards potentiels. De plus, les limites de la consommation électrique des appareils peuvent créer des contraintes sur la fréquence à laquelle une transmission de paquets peut avoir lieu. Étant donné que les performances de tout système de mise à jour d'état dépendent fortement de la livraison *opportune* de ces paquets d'un bout à l'autre, un cadre d'optimisation du système adéquat doit être adopté pour atteindre cet objectif. Les deux mesures standard utilisées pour optimiser les réseaux de communication sont le débit et le délai. Le débit est défini comme la quantité de données envoyées par la source dans une unité de temps. D'autre part, le délai est le temps mis par un paquet pour se déplacer de la source au moniteur distant. Sur la base de ces métriques, deux cadres d'optimisation émergent:

- **Maximization de Débit**: Comme son nom l'indique, ce cadre vise à maximiser la quantité de données générées et envoyées par la source. Cela se fait en augmentant le taux de génération de paquets et en utilisant le canal dans la plus grande mesure possible. Cette approche ne parvient cependant pas à atteindre l'objectif du système de mise à jour d'état. Plus précisément, en raison de l'augmentation du taux de génération de paquets, des délais de mise en

file d'attente élevés seront encourus. Par conséquent, le moniteur recevra des paquets obsolètes pour lesquels le temps de livraison est nettement supérieur à leur temps de génération.

- **Minimisation des Délais**: Ce cadre vise à réduire le temps nécessaire pour qu'un paquet passe de l'émetteur au récepteur. Ceci est réalisé en évitant les délais de mise en file d'attente et en utilisant le canal de sorte qu'aucun délai supplémentaire ne soit introduit. Naturellement, cela conduira à une diminution du taux de génération de paquets afin de réduire la charge sur le système. En fait, du point de vue du délai de bout en bout, "*il n'y a pas de délai s'il n'y a pas de transmission*". En raison de la faible génération de paquets, le moniteur distant ne pourra pas être mis à jour adéquatement.

Une question importante est donc la suivante: quelle est la métrique alternative que nous devrions adopter pour optimiser le système de mise à jour d'état? En réponse à cette question, la notion de l'Âge de l'Information (AdI) est née.

1.2 Âge de l'Information

Pour comprendre la métrique de l'AdI, considérons le système de mise à jour d'état rapporté dans la Fig. 1.2. Le système commence à fonctionner au temps $t = 0$. Le paquet de mise à jour d'indice i est généré au temps S_i, arrive à la file d'attente au temps A_i et est delivré au moniteur au temps D_i. En conséquence, nous avons toujours $0 \leq S_1 \leq S_2 \leq \ldots$ et $S_i \leq A_i \leq D_i$. L'AdI à l'instant t est défini comme [8]:

$$\Delta(t) = t - \max\{S_i : D_i \leq t\}, \tag{1.1}$$

qui est la différence entre le temps courant t et le temps de génération du paquet le plus frais qui a été livré au moniteur. De cette définition, on peut voir qu'une petite valeur de $\Delta(t)$ implique qu'il existe un nouvel échantillon d'état au moniteur. En revanche, un âge élevé signifie que le moniteur n'a pas été mis à jour sur le processus d'intérêt pendant une longue durée. Par conséquent, en minimisant la valeur de l'AdI, nous pouvons garantir que le moniteur dispose de nouvelles données sur le processus observé par la source. Comme le montre la Fig. 1.3, $\Delta(t)$ évolue comme une courbe en dents de scie où des chuttes de la valeur d'âge se produisent lorsqu'un nouvel échantillon est reçu par le moniteur. Nous laissons $\{\Delta(t), t \geq 0\}$ désigner le processus d'âge, et nous définissons C_i comme la i-ème valeur d'âge de pointe du processus d'âge $\{\Delta(t), t \geq 0\}$ dès $t = 0$.

Le concept d'âge remonte aux années 1990 où il a été introduit dans le cadre du maintien de la fraîcheur des bases de données en temps réel [9]. Grâce à l'essor récent des applications IdO en temps réel, ce concept a été réintroduit du point de vue du réseau de files d'attente dans [8]. Plus précisément, la minimisation de l'AdI a été étudiée pour les modèles de file d'attente First-Come-First-Served (FCFS) standard: M/M/1, M/D/1 et D/M/1. Les auteurs ont pu montrer l'existence d'une fréquence optimale à laquelle la source doit générer ses mises à jour d'état

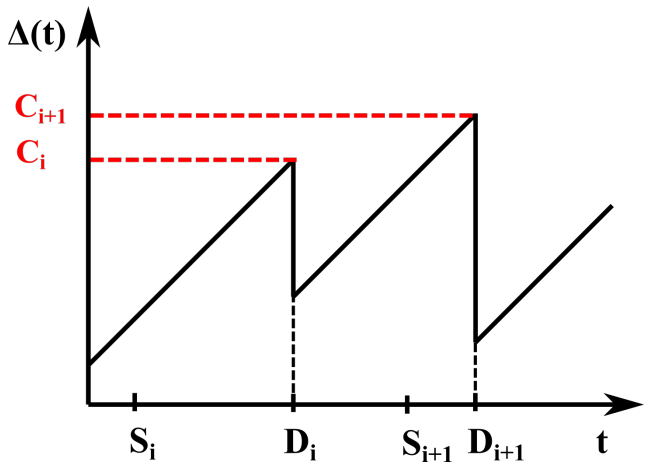


Figure 1.3: Illustration de l'évolution de l'AdI.

pour maintenir le moniteur à jour dans les meilleurs délais. Ce taux s'est avéré *différent* de ceux qui maximisent le débit ou minimisent le délai de livraison des paquets. Les résultats de ce travail fondateur ont suscité l'intérêt de la communauté de recherche, et une augmentation du nombre d'articles publiés sur ce sujet peut être observée [10, 11].

Au cœur de tous les travaux de recherche dans ce domaine, des métriques basées sur l'âge ont été utilisées pour représenter le niveau d'insatisfaction d'avoir des informations âgées chez le moniteur. Ces métriques sont des fonctionnelles du processus d'âge $\{\Delta(t), t \geq 0\}$. Parmi les métriques d'âge largement utilisées, nous citons les suivantes:

- La moyenne temporelle de l'âge [8]:

$$\phi_{\text{avg}}(\{\Delta(t), t \geq 0\}) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T \Delta(t) dt. \tag{1.2}$$

- L'âge maximal moyen Maximum Âge de l'Information (MAdI) [12]:

$$\phi_{\text{peak}}(\{\Delta(t), t \geq 0\}) = \lim_{K \to +\infty} \frac{1}{K} \sum_{i=1}^{K} C_i, \tag{1.3}$$

  où C_i est la i-ème valeur maximal de l'âge depuis $t = 0$ comme le montre la Fig. 1.3.

- La moyenne temporelle d'une fonction non décroissante de l'âge p_t [7]:

$$\phi_{\text{avg-pen}}(\{\Delta(t), t \geq 0\}) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T p_t(\Delta(t)) dt, \tag{1.4}$$

  où $p_t : [0, +\infty[\mapsto \mathbb{R}$ est une fonction non décroissante. La propriété de non décroissance est naturelle et est conforme au fait que des données fraîches

sont souvent plus recherchées que des données obsolètes. Il a été démontré que les fonctions d'âge non-linéaires sont étroitement liées à de nombreuses applications en temps réel où diverses mesures de performance peuvent être converties en moyenne temporelle d'une fonction non décroissante de l'âge. Parmi ces grandeurs, nous citons l'erreur d'estimation et l'auto-corrélation des signaux temps réel [13].

Équipés de ces mesures d'âge, les chercheurs ont examiné les potentiels de l'AdI.

1.3 État de l'Art et Tendances de la Recherche

Les avantages de l'AdI dans la modélisation de diverses applications réelles sont régulièrement mis en lumière, mettant en évidence la large portée de cette métrique. Par exemple, des études approfondies sur le problème de planification de trajectoire pour les réseaux des véhicules aériens sans pilote ont été réalisées dans la littérature sur la base de nombreux principes de conception bien connus tels que le débit, l'efficacité énergétique et le temps de vol. Les auteurs dans [14] ont développé une planification de chemin basée sur l'âge pour ces réseaux et ont mis en évidence les gains de fraîcheur des informations par rapport aux conceptions d'algorithmes de planification de chemin traditionnels. Dans une autre ligne de travail, les problèmes d'estimation de canal dans les réseaux sans-fil ont été étudiés [15]. En particulier, dans de nombreux systèmes de communication sans-fil, l'émetteur s'adapte à l'état instantané du canal qui varie avec le temps en utilisant une estimation de canal à sa disposition. Les émetteurs peuvent utiliser des informations obsolètes sur l'état des canaux en raison du coût élevé de l'entraînement. Les auteurs dans [15] ont pu caractériser la dégradation des performances due à ce phénomène en se basant sur le concept de l'âge de l'information. Récemment, l'utilisation potentielle de l'AdI pour maintenir à jour l'index des citations des chercheurs sur des sites Web tels que Google Scholar a été étudiée dans [16]. Comme on peut le voir, la portée des applications que l'AdI englobe est vaste et le terrain est fertile pour de nombreuses autres à venir. Cela rend l'analyse des métriques basées sur l'âge dans des environnements généraux essentielle pour avoir une meilleure compréhension de la métrique. Cela a été un axe fondamental du travail de recherche dans le domaine, et c'est l'un des principaux objectifs que cette thèse vise à atteindre. Parmi ces travaux de recherche, les plus pertinents pour cette thèse peuvent être classés en trois grandes catégories:

1. Optimisation de l'âge des systèmes de files d'attente
2. Planification en utilisant l'âge
3. Applications d'estimation à distance

1.3.1 Optimisation de l'Âge des Systèmes de Files d'Attente

Comme pour les métriques de débit et de délai, la première étape pour mieux comprendre une nouvelle métrique et sa dynamique est de l'analyser et de l'optimiser

dans les systèmes de mise en file d'attente généraux. Depuis sa renaissance dans [8], l'AdI a été examiné dans un large éventail de systèmes de files d'attente. Dans [12], les auteurs ont montré que la gestion des paquets peut réduire l'AdI par rapport à la discipline FCFS. En suivant les mêmes traces, la politique Last-Generated-First-Served (LGFS) s'est avérée optimale pour minimiser l'AdI dans les réseaux à sauts simples et multiples [17]. Les avantages d'avoir des serveurs parallèles sur les performances de l'AdI ont été explorés dans [18]. L'effet des pertes de paquets sur l'AdI a été analysé dans [19]. Des méthodes de codage de source et de codage de canal pour minimiser l'âge ont également été proposés dans [20–22].

Les systèmes de mises à jour d'état où la source est contrainte par la consommation d'énergie ont été fortement étudiés [23–26]. Par exemple, les auteurs dans [23] ont étudié le cas où le taux de mise à jour de la source ne peut pas dépasser une limite prédéfinie en raison de limites de batterie. Les systèmes de récupération d'énergie ont également attiré une part équitable de l'attention de la recherche. Dans [25], Yates a examiné le cas où un système de récupération d'énergie stochastique alimente la source. Il a montré que la politique de l'âge optimal est paresseuse. En d'autres termes, après l'achèvement d'un service, le serveur est fréquemment laissée inactive même si le serveur a suffisamment d'énergie pour soumettre un paquet de mise à jour. Les résultats diffusés par Yates dans [25] étaient surprenants car on pensait qu'un nouveau paquet devrait être envoyé chaque fois que les niveaux d'énergie disponibles le permettent. Capitalisant sur ces résultats, Sun et al. ont étudié un canal de délai où la source peut générer des paquets à volonté [7]. Le système est rapporté dans la Fig. 1.4. Ils ont pu montrer des résultats similaires selon lesquels une politique d'échantillonnage sans attente peut être loin d'être optimale pour l'âge. Dans la politique d'échantillonnage sans attente, un nouvel échantillon est généré et transmis une fois le paquet de mise à jour précédent est livré, et un paquet Acknowledgement (ACK) est reçu. Les résultats dans [7] suggèrent que la source devra peut-être attendre avant de générer et de transmettre le paquet. Le temps d'attente dépend de la distribution du délai du canal et de l'âge actuel chez le moniteur. Des résultats similaires ont été trouvés plus tard pour le cas des arrivées stochastiques dans [27].

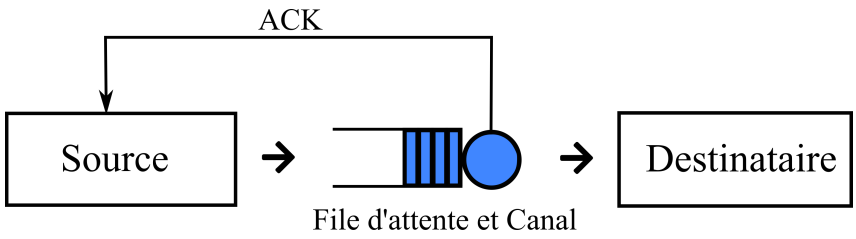


Figure 1.4: Modèle de génération à volonté avec ACK.

Il existe également plusieurs études récentes sur les mises à jour d'état avec plusieurs classes de priorité. C'est une conséquence naturelle de la grande quantité de scénarios réels où les flux d'informations se voient attribuer des priorités différentes en fonction de leur sensibilité. Dans [28], les auteurs ont considéré

plusieurs flux d'informations, chacun avec une priorité différente, et partageant un serveur commun sans et avec une salle d'attente dans la file d'attente partagée par les flux. Ils ont étudié le cas où un paquet de haute priorité préempterait un paquet de priorité inférieure, qui est ensuite rejeté. Ils ont dérivé une expression de l'âge moyen de chaque flux, et le taux d'arrivée de chaque flux a ensuite été optimisé en conséquence. Plus récemment, des expressions de l'âge maximal moyen ont été trouvées dans les systèmes M/M/1/1, où les flux sont attribué des priorités différentes [29].

En raison de l'importance des environnements multi-classes, nous examinons dans le Chapitre 3 de cette thèse un système de files d'attente avec priorité. Nous analysons l'âge moyen du système et fournissons des résultats sur l'interaction entre les différentes classes. Ces résultats constitueront l'épine dorsale de notre proposition d'un nouveau cadre d'optimisation de l'âge pour les systèmes multi-classes.

1.3.2 Planification en Utilisant l'Âge

Dans la majorité des applications en temps réel, les flux d'informations concurrents partagent le canal de transmission où les ressources disponibles peuvent être rares. La rareté peut être le résultat de considérations de batterie pour les dispositifs impliqués ou d'interférences physiques limitant le nombre de transmissions simultanées entre les flux d'informations. En conséquence, une affectation adéquate des ressources disponibles pour atteindre l'objectif désiré doit être adoptée. À cette fin, les chercheurs ont largement exploré une gamme de problèmes de planification visant à minimiser les métriques basées sur l'âge dans divers contextes aux ressources limitées. Les efforts peuvent être classés en deux parties principales.

Planification Centralisée

Dans la planification *centralisée*, un seul planificateur dédié gère la gestion et l'allocation des ressources disponibles. Le planificateur dispose généralement d'informations précises sur l'état de l'ensemble du système. Les informations disponibles à sa disposition couvrent 1) la topologie du réseau, 2) l'AdI instantané chez le moniteur, 3) les bilans de consommation de batterie de chaque appareil et 4) les conditions de canal. De nombreux réseaux appartiennent à cette catégorie; peut-être le plus connu est le réseau cellulaire standard où la station de base gère la planification des périphériques dans la cellule.

Jusqu'à présent, les chercheurs ont examiné les métriques basées sur l'AdI dans une variété de systèmes. Dans [30], les auteurs ont étudié l'âge moyen pondéré dans un environnement multi-utilisateurs où, au plus, un utilisateur peut transmettre et les canaux présentent des erreurs de décodage possibles. Il a été démontré qu'un algorithme glouton est optimal lorsque les utilisateurs ont des statistiques de canal identiques. Dans le cas asymétrique, des politiques sous-optimales ont été proposées. De même, une approche d'approximation Markov Decision Process (MDP) a été utilisée pour proposer des politiques de planification à la fois hors ligne

et en ligne dans le cas asymétrique [31]. Dans les mêmes contextes mais avec des arrivées aléatoires, une politique d'index de Whittle a été proposée pour minimiser l'âge moyen [32]. Dans [33], les auteurs ont montré que sous l'hypothèse des temps d'arrivée et de génération des paquets synchronisés, un algorithme glouton est optimal dans le scénario d'un seul serveur exponentiel. Dans une autre ligne de travail, le modèle physique Signal-to-Interference-plus-Noise Ratio (SINR) a été examiné dans un contexte multi-utilisateurs. Il a été démontré que trouver la politique d'âge moyen minimum est NP-difficile, et une heuristique de faible complexité a donc été proposée. Dans le cas des canaux de délai [34], il a été prouvé que la politique d'échantillonnage sans attente minimise l'âge maximal moyen dans les systèmes multi-utilisateurs.

À ce jour, de nouveaux travaux de recherche sur la planification en fonction de l'âge dans des environnements centralisés sont en cours de publication. Nous encourageons les lecteurs à se référer à [35] pour la liste des publications les plus récentes sur le sujet. Comme de nombreux systèmes sont distribués par nature et que les méthodes centralisés ne peuvent donc pas être adoptés, des algorithmes d'ordonnancement distribués ont été étudiés dans la littérature.

Planification distribuée

Dans de nombreux scénarios, en particulier les applications IdO, il n'est pas nécessairement possible de disposer d'une entité centrale qui gère les ressources disponibles. Cette infaisabilité est causée par diverses raisons telles que le coût, la complexité, les considérations de puissance et la nature autonome des appareils. Par conséquent, au lieu d'une approche centralisée, une approche distribuée doit être adoptée où les appareils n'utilisent que les informations locales à leur disposition pour accéder aux ressources disponibles. L'AdI étant d'un grand intérêt dans ce type d'application, des efforts de recherche considérables ont été consacrés à l'optimisation de l'AdI dans des environnements de planification décentralisés.

Parmi ces efforts, les auteurs dans [36] ont étudié une méthode d'ordonnancement distribuée où les appareils qui disposent toujours d'informations actualisées accèdent au canal avec une certaine probabilité. La probabilité d'accès de chaque lien est optimisée pour minimiser l'âge moyen total du réseau. Dans [37], il a été montré qu'un algorithm d'accès multiple Round Robin (RR) distribué atteint l'âge moyen minimum lorsque le nombre d'appareils tend vers l'infini pour les canaux sans bruit. Dans le cas des canaux bruyants, les auteurs dans [38] ont proposé une méthode d'accès décentralisée basée sur l'indice de Whittle. Il a été démontré numériquement qu'elle possédait de bonnes performances dans le régime de nombreux appareils. Dans une autre ligne de travail, la méthode d'accès au canal Carrier Sense Multiple Access (CSMA) a été étudié dans le cadre des réseaux véhiculaires dans [39]. Dans ce travail, les auteurs ont étudié l'âge moyen d'un environnement de réseau véhiculaire *numériquement*. Il a été montré que l'âge de l'information est minimisé à un point de fonctionnement optimal entre les deux extrêmes du débit maximal et du délai minimal.

Bien que le CSMA est considéré comme l'un des méthodes d'ordonnancement distribués les plus connues et les plus répandues, la caractérisation théorique de son point de fonctionnement optimal d'âge reste une question ouverte. Pour y remédier, nous étudions théoriquement dans Chapitre 4 un environnement CSMA général et fournissons les premiers résultats théoriques sur son point de fonctionnement qui minimise l'AdI moyen.

1.3.3 Applications d'Estimation à Distance

Parmi les diverses applications de l'AdI que les chercheurs ont étudiées, peut-être celles qui ressortent le plus sont les problèmes d'estimation à distance. La raison principale en est la large applicabilité de ces problèmes dans des scénarios réels. Dans les applications d'estimation à distance, la source observe un processus X_t qui change avec le temps. X_t peut être, par exemple, la température d'une pièce, la vitesse d'un véhicule ou les commandes d'un contrôleur. En utilisant les paquets de mise à jour d'état envoyées par la source, le moniteur construit une estimation de X_t, notée $\hat{X}_t$. Le but est de rendre $\hat{X}_t$ aussi proche que possible de X_t. L'approche standard pour atteindre cet objectif est de minimiser l'erreur d'estimation $\Pr(X_t \neq \hat{X}_t)$ ou le Mean Squared Error (MSE) $\mathbb{E}[|X_t - \hat{X}_t|^2]$. Les chercheurs ont montré, dans de nombreux travaux, que l'AdI est étroitement liée aux deux objectifs mentionnés ci-dessus. Par exemple, Sun et al. ont étudié la politique Minimum Mean Squared Error (MMSE) dans un canal de délai où X_t est un processus Wiener [40]. Ils ont montré que trouver la politique d'échantillonnage MMSE équivaut à trouver la politique d'échantillonnage qui minimise la moyenne d'une fonction non linéaire de l'âge. Leurs résultats sont valables pour le cas où les instants d'échantillonnage sont indépendants des valeurs de X_t. De même, dans Network Controlled Systems (NCS) où l'installation et le contrôleur sont régis par un système Linear Time-Invariant (LTI), la minimisation d'une fonction non linéaire particulière de l'AdI conduit à la minimisation du MSE [41]. Dans une autre ligne de travail, la maximisation de l'information mutuelle entre les deux processus X_t et $\hat{X}_t$ lorsque X_t est un processus Markovien stationnaire s'est également avérée liée à la minimisation de la moyenne d'une fonction non-linéaire de l'âge [42].

Dans le même temps, les lacunes de l'AdI ont également été mises en perspective pour des problèmes particuliers d'estimation en temps réel. Par exemple, lorsque X_t est un processus Wiener et que les temps d'échantillonnage peuvent dépendre de X_t, la politique MMSE dans un canal de délai n'est pas équivalente à une minimisation d'une métrique basée sur l'âge [40]. De même, l'AdI a été montré dans [43] comme sous-optimal en minimisant l'erreur d'estimation sur les canaux avec des erreurs de décodage potentielles lorsque X_t est un processus de Markov. La principale explication de ces défauts est que l'AdI, par définition, ne dépend pas de X_t et $\hat{X}_t$. Ces observations ont incité à proposer de nouvelles mesures de performance qui traitent les lacunes de l'AdI. Parmi ces efforts, une métrique temporelle surnommée l'Âge de Synchronisation (AdS) a été introduite dans le cadre de la mise à jour des caches [44, 45]. Plus précisément, l'AdS est nul lorsque l'émetteur n'a

aucun paquet à envoyer. Il croît linéairement avec le temps lorsque le côté émetteur génère un nouveau paquet. Bien que l'AdS inclue la génération de paquets comme un facteur, cela ne dépend pas de X_t et $\hat{X}_t$, ce qui peut limiter son utilisation dans les applications d'estimation à distance. Dans un autre travail [46], les auteurs ont proposé différentes métriques d'âge *effectif* pour lesquelles un faible âge effectif devrait sans aucun doute conduire à une erreur de prédiction plus faible. Par exemple, la notion de l'*âge d'échantillonnage* a été introduite, qui a été définie comme l'âge par rapport à un modèle d'échantillonnage idéal $g(t)$ qui minimise l'erreur. Cependant, trouver le modèle optimal $g(t)$ a été jugé loin d'être trivial.

Dans le Chapitre 5 de cette thèse, en plus des lacunes de l'AdI, nous mettons en lumière les lacunes des fonctions de pénalité d'erreur conventionnelles (par exemple, erreur d'estimation, erreur quadratique moyenne). En conséquence, nous proposons une nouvelle métrique de performance qui traite ces lacunes et nous présentons ses avantages dans de nombreuses applications.

1.4 Aperçu de la Thèse et Contributions

Cette thèse est divisée en deux parties. Dans la partie I, nous nous concentrons sur l'optimisation des mesures basées sur l'âge dans divers systèmes de communication. Plus précisément:

Dans le Chapitre 3 (Multi-Class Multi-Stream Scheduling), nous considérons un problème de planification, dans lequel plusieurs flux de paquets de mise à jour d'état avec divers niveaux de priorité sont envoyés via un canal partagé vers leurs destinations. Nous trouvons une expression de l'âge moyen de chaque flux pour le cas où les arrivées suivent une loi de Poisson et les temps de transmission sont exponentiels. En utilisant ces expressions, nous fournissons des résultats sur l'interaction entre les différentes classes. Ces résultats suggèrent la nécessité d'un nouveau cadre d'optimisation de l'âge dans les systèmes multi-classes. Pour cela, nous introduisons une notion d'optimalité d'âge lexicographique, ou simplement l'optimalité lex-âge, pour évaluer la performance des politiques de mise à jour d'état. En particulier, une politique est dite lex-âge-optimal si elle minimise d'abord les métriques de l'AdI pour les flux hautement prioritaires, puis, dans l'ensemble des politiques optimales pour les flux hautement prioritaires, elle optimise les métriques de l'AdI pour les flux à faible priorité. Nous proposerons une nouvelle politique d'ordonnancement nommée Preemptive-Priority, Maximum-Age-First,Last-Generated First-Served (PP-MAF-LGFS), et nous prouverons que cette politique est lex-âge-optimal. Ce résultat est valable (i) pour minimiser tout fonction de pénalité d'âge symétrique et non décroissante; (ii) pour minimiser toute fonctionnelle non-décroissante du processus stochastique formé par la fonction pénalité d'âge; et (iii) pour les cas où les différentes classes de priorité ont des modèles de trafic et fonctions de pénalité d'âge distinctes.

Dans le Chapitre 4 (Average Age Minimization in a CSMA Environment), nous étudions un réseau où N utilisateurs accèdent au canal en utilisant la méth-

ode d'accés CSMA. CSMA est une classe de protocoles d'accès multiple simple et distribué et qui est considéré comme l'un des protocole le plus répandu dans les réseaux sans-fil (par exemple, CSMA est le algorithme d'accès de base dans IEEE 802.11 [47]). Dans cette classe de protocoles, l'émetteur attend pour une certain durée de temps avant de transmettre, appelé le back-off temps. Dans cet environnement de transmission, nous dérivons des expressions de l'âge moyen en fonction des paramètres du système. Equipé de ces expressions, nous formulons le problème de la minimisation de l'âge moyen en calibrant le temps de back-off de chaque lien. En analysant sa structure, nous convertissons le problème d'optimisation formulé à un problème convexe équivalent dont on dérive sa solution optimale. Un aperçu de l'interaction entre les liens et des implémentations numériques de cette méthode optimisé dans un environnement IEEE 802.11 sont également présentés. De plus, pour améliorer la performance de cette méthode, nous proposons une version modifiée du protocole CSMA en donnant à chaque lien la liberté de passer en mode SLEEP. L'approche proposée offre un moyen de réduire la charge pesant sur le canal quand c'est possible. Cela conduit à une amélioration des performances du réseau.

Dans la partie II de la thèse, nous proposons une nouvelle mesure de performance qui addresse les lacunes de l'AdI et des fonctions d'erreur conventionnelles largement utilisées. En particulier:

Dans le Chapitre 5 (Age of Incorrect Information: Analysis and Optimization), nous introduisons une nouvelle métrique dans le cadre des systèmes de mise à jour d'état que nous appelons l'Âge de l'Information Incorrecte (AdII). Cette nouvelle métrique traite les lacunes de l'AdI et des fonctions d'erreur conventionnelles comme elle étend proprement la notion de paquets de mise à jour à celle de paquets «informatifs» de mise à jour. Le mot informatif dans ce contexte se réfère aux mises à jour qui apportent des informations nouvelles et correctes au moniteur. L'AdII capture également l'effet de détérioration que les informations erronées peuvent avoir avec le temps sur le système. Après avoir motivé la nouvelle métrique, nous trouvons la politique d'échantillonnage optimale qui minimise l'AdII dans les scénarios sans contrainte et avec contrainte de ressources. Ces politiques sont ensuite mises en œuvre et leurs avantages par rapport aux politiques optimales pour l'âge et les fonctions d'erreur dans une variété d'applications sont soulignés.

Enfin, nous présentons nos conclusions et travaux futurs qui pourront être entrepris dans le prolongement de cette thèse. Nous notons que chaque chapitre ci-dessus contient ses propres notations mathématiques.

1.5 Publications

Les publications suivantes ont été produites au cours de cette thèse [48–63]. Les résultats qui sont entièrement ou partiellement fournis dans cette thèse sont marqués par *.

Article dans une Revue

[48]* 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, pp. 1–14, 2020.

[49]* A. Maatouk, M. Assaad, and A. Ephremides, “On The Age of Information in a CSMA Environment," IEEE/ACM Transactions on Networking, vol. 28, no. 2, pp. 818–831, 2020.

[50] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “On The Optimality of The Whittle’s Index Policy For Minimizing The Age of Information," revisions stages in IEEE Transactions on Wireless Communications.

[51] A. Maatouk, S. E. Hajri, M. Assaad, and H. Sari, “On Optimal Scheduling for Joint Spatial Division and Multiplexing Approach in FDD Massive MIMO," IEEE Transactions on Signal Processing, vol. 67, no. 4, pp. 1006–1021, 2019.

[52] A. Maatouk, M. Assaad, and A. Ephremides, “Energy Efficient and Throughput Optimal CSMA Scheme," IEEE/ACM Transactions on Networking, vol. 27, no. 1, pp. 316–329, 2019.

[53] A. Maatouk, E. Caliskan, M. Assaad, M. Koca, G. Gui, and H. Sari, “Frequency-Domain NOMA With Two Sets of Orthogonal Signal Waveforms," IEEE Communications Letters, vol. 22, no. 5, pp. 906–909, 2018.

Communication dans un Congès

[54]* 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), pp. 1–8, 2020.

[55]* A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides “Asymptotically Optimal Scheduling Policy For Minimizing The Age of Information," in 2020 IEEE International Symposium on Information Theory (ISIT), pp. 1747–1752, 2020.

[56]* 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), pp. 1– 8, 2019.

[57] 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), pp. 102–108, 2019.

[58]* 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), pp. 325–329, July 2019.

[59] A. Maatouk, M. Assaad and A. Ephremides, “The Age of Updates in a Simple Relay Network," in 2018 IEEE Information Theory Workshop (ITW), pp. 1–5, 2018.

[60] E. Caliskan, A. Maatouk, M. Koca, M. Assaad, G. Gui and H. Sari, “A Simple NOMA Scheme with Optimum Detection," in 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2018.

[61] J. Denis, A. Maatouk, S. E. Hajri and M. Assaad, “Stay Longer at the Network’s Edge: A Novel Proactive Caching Policy through Sojourn Time," in 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2018.

[62] A. Maatouk, S. E. Hajri, M. Assaad, H. Sari and S. Sezginer, “Graph Theory Based Approach to Users Grouping and Downlink Scheduling in FDD Massive MIMO," in 2018 IEEE International Conference on Communications (ICC), pp. 1–7, 2018.

[63] H. Sari, A. Maatouk, E. Caliskan, M. Assaad, M. Koca and G. Gui, “On the foundation of NOMA and its application to 5G cellular networks," in 2018 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1– 6, 2018.

2 | Introduction

2.1 Background and Motivation

By nature, human beings are a social species; we share mirror neurons that enable us to match each other's emotions immediately and subconsciously. We even mirror each other's brain activity when we are engaged in storytelling. We create bonds, express emotions, and form societies. In his famous book "Politics", Aristotle described humans as follows:

Man is by nature a social animal.

These fundamental observations point to the fact that none can break the shackles of mutual dependence. Humans must satisfy their natural social needs and have their sense of belonging. Let it be his/her home, community, or state; a social environment is mandatory to the stability of an individual. Sociologists consider that this starts in the relationship between the embryo and the mother, and continues till death.

At the core of our social interactions lays the fundamental concept of *communication*. The root of the word communication is the Latin word *communicare*, which means sharing or making common. To that end, communication is defined as a process between at least two entities where a transfer of information from one entity to the other takes place. Communication is the backbone of all human relationships. It helps us spread knowledge, and allows us to express our ideas and feelings and understand others' emotions and thoughts. Correspondingly, affection is developed, and both positive and negative relationships can be founded. Therefore, we cannot underestimate the importance of communication. In its simplest form, and as seen in Fig. 2.1, any communication process has three essential components:

1. **Source**: The source is the entity that creates and sends the message. He or she begins by first deciding on the message desired to be communicated. The message is then encoded by determining the perfect combination of verbal (e.g., words) and non-verbal (e.g., body language, voice tone) expressions that convey the intended meaning. Finally, the message is propagated through a medium in order to reach the sought-after audience.

2. **Channel**: The channel is the means in which the encoded message travel between the source and receiver. For example, in spoken situations like face-to-face conversations, the medium is the air in which mechanical sound waves propagate. The channel is ordinarily subject to several perturbations that may hinder the communication process. For instance, in the same face-to-face conversation, the noise from the wind, or the voices of other speakers may affect the communication process at hand.

3. **Receiver**: The receiver is the entity for whom the message is intended. After traveling through the channel, the message is decoded and analyzed by the receiver. The degree to which the receiver understands the message is heavily reliant on numerous factors. Among these factors, we cite the noise and the interference in the channel, and the message's reception time.

Figure 2.1: Basic communication process.

Driven by their social needs, humans kept pioneering new forms of communication beyond the standard face-to-face conversations. From the use of hieroglyphic writing in ancient Egypt to the invention of the alphabet by the Phoenicians, human innovation in this regard never stopped. In the last couple of centuries, the advances in technology have created a fertile environment from which emerged a connected world where communication from almost any point of the earth (and beyond) to the other is both feasible and fast. From satellites broadcasting to smartphones' proliferation, the world has never been as connected as it is today. All these recent advances in communication systems have radically changed how humans communicate and access/exchange information. This radical change reached a point where communication surpassed the human-to-human communication exclusivity. Specifically, thanks to the ubiquitous connectivity and the cheap hardware cost, the Internet of Things (IoT) was born.

In the broadest sense, an IoT system consists of interrelated computing devices, mechanical and digital machines that can transfer data over a network without requiring human-to-human or human-to-computer interaction [1]. In their 2013 report, McKinsey notes a 300% growth in connected IoT devices in the last five years and rate the potential economic impact of the IoT at $2.7 trillion to $6.2 trillion annually by 2025 [2]. This showcases the momentum IoT systems are gaining in both the industry and research communities and its importance in our future daily lives. We report below various examples of IoT systems applications.

- **Vehicular Networks**: These networks are formed among moving vehicles, Road Side Units (RSUs), and pedestrians that carry communication devices [3]. The messages generated and sent by these entities throughout the network

contain data such as vehicles' velocity and position, traffic lights, accidents updates, etc. [4].

- **Smart Homes**: In smart homes applications, an individual can remotely manage, monitor, and control various building services such as climate, entertainment systems, lighting, and home appliances [5].

- **Environmental Monitoring**: This branch of applications covers processes and activities necessary to identify and monitor the quality of the environment (e.g., air, water, and soil qualities). Thanks to the recent advances in hardware and wireless technologies, this can be achieved by relaying data from sensors distributed in the environmental area of interest to a remote monitor for processing [6].

At the core, the above applications and many other IoT applications, belong to the same category of IoT services called "**Real-time Monitoring**". In such systems, an entity is interested in knowing the status of one or multiple processes observed by a remote source. Accordingly, the source sends *status updates* packets to the monitor to provide information about the process/processes of interest. We distinguish in this case between two scenarios:

- **Generate-at-will**: In this scenario, the remote source can generate status updates at any desired time instance [7]. An example is when the source is a temperature sensor aiming to reproduce the temperature process at the monitor side.

- **Stochastic arrivals**: In this scenario, status updates arrive randomly to the source according to a probability distribution. This is the scenario where the remote source has no control over the packets' arrivals.

In both scenarios, once the status update packet arrives at the buffer, it is prepared for transmission over the channel. Afterward, the update is sent through the network, which can be constituted of multiple wired/wireless links. Once the status update is delivered to the monitor, it can proceed with performing its tasks. This status updates system is reported in Fig. 2.2. We note that the status updates system at hand is nothing but a special case of the basic communication model depicted in Fig. 2.1. Specifically, in the status updates system, the source/receiver can be either machines or humans, and the channel is an arbitrary combination of both wired and wireless links.

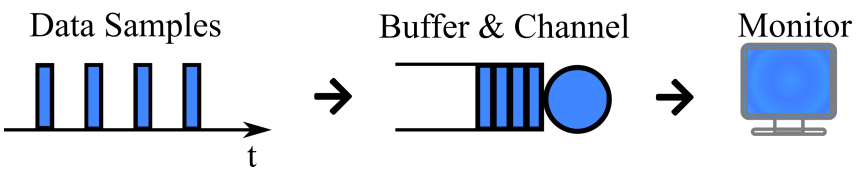


Figure 2.2: Status updates system.

The main objective in these applications is to keep the monitor informed and up to date on the observed process by the source. In other words, information has the highest value when it is fresh. The older a packet is, the lower the value it brings to the monitor. To understand this better, let us take the example of a vehicular network where vehicles' position is disseminated to a central entity that manages the traffic lights. To reduce traffic jams and to make traffic flow more fluid, the central entity controls traffic lights based on the information it has about the various vehicles. Clearly, the efficiency of the controller's decisions will be heavily reliant on how up to date it is on the vehicles' positions. For instance, suppose that the controller has information about vehicles' position 15 minutes ago. The algorithm used to optimize traffic will not lead to the same results compared to having information about the vehicles' position just 5 seconds ago. The algorithm will probably make the traffic worse due to the use of outdated data. This observation puts into perspective how reliant these applications are on fresh data. With that in mind, we note that from generation till reception, the status update packet can be subject to queuing congestion, channel errors, and other potential delays. Moreover, limits on the power consumption of devices can create constraints on how frequently a packet transmission can occur. As the performance of any status updates system is heavily dependent on the *timely* delivery of these packets from one end to the other, an adequate system optimization framework needs to be adopted to achieve the timeliness goal. The two standard metrics employed for optimizing communication networks are throughput and delay. Throughput is defined as the amount of data sent by the source within a unit of time. On the other hand, delay is the time taken by a packet to travel from the source to the remote monitor. Based on these metrics, two optimization frameworks emerge:

- **Throughput maximization**: As the name suggests, this framework aims to maximize the amount of data generated and sent by the source. This is done by increasing the generation rate of packets and utilizing the channel to the greatest extent possible. This approach, however, falls short in achieving the timeliness goal for the status updates system. Specifically, due to the increased generation rate of packets, high queuing delays will be incurred. Consequently, the monitor will be receiving stale packets for which the delivery time is significantly larger than their generation time.

- **Delay minimization**: This framework seeks to reduce the time necessary for a packet to go from the transmitter to the receiver. This is achieved by avoiding queuing delays and utilizing the channel so that no additional delays are introduced. Naturally, this will lead to a decrease in the generation rate of packets in order to reduce the burden on the system. In fact, from an end-to-end delay perspective, "*there is no delay if there is no transmission*". Because of the scarce generation of packets, the remote monitor will be unable to be kept updated in a timely fashion.

An important question, therefore, is: What is the alternative metric that we should

adopt to optimize the status updates system? As an answer to this question, the notion of the Age of Information (AoI) was born.

2.2 Age of Information

To understand the AoI metric, let us consider the status updates system reported in Fig. 2.2. The system starts operating at time $t = 0$. The i-th status update packet is generated at time S_i, arrives to the buffer at time A_i and is delivered to the monitor at time D_i. Accordingly, we always have $0 \leq S_1 \leq S_2 \leq \ldots$ and $S_i \leq A_i \leq D_i$. The instantaneous AoI at time instant t is defined as [8]:

$$\Delta(t) = t - \max\{S_i : D_i \leq t\}, \tag{2.1}$$

which is the difference between the current time t and the generation time of the freshest packet that has been delivered to the monitor. From this definition, one can see that a small value of $\Delta(t)$ implies that there exists a fresh status sample at the monitor. In contrast, a large age signifies that the monitor has not been updated on the process of interest for a large duration of time. Therefore, by minimizing the value of the AoI, we can guarantee that the monitor has at its disposal fresh data about the observed process by the source. As plotted in Fig. 2.3, $\Delta(t)$ evolves as a sawtooth curve where drops in the age value happen when a new sample is received by the monitor. We let $\{\Delta(t), t \geq 0\}$ denote the age process, and we define C_i as the i-th peak age value of the age process $\{\Delta(t), t \geq 0\}$ since $t = 0$.

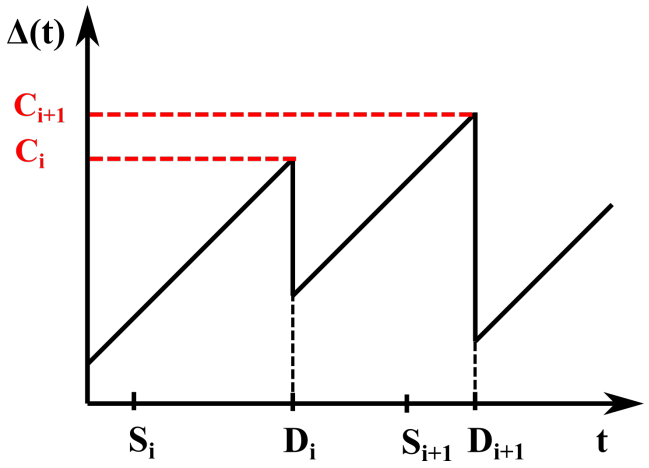


Figure 2.3: Illustration of the AoI evolution.

The concept of age dates back to as early as the 1990's where it was introduced in the framework of keeping the freshness of real-time databases [9]. Thanks to the recent boom in real-time IoT applications, this concept was reintroduced from a queuing network perspective in [8]. Specifically, the minimization of the AoI was investigated for the standard First-Come-First-Served (FCFS) queuing models: M/M/1, M/D/1 and D/M/1. The authors were able to show the existence of an

optimal rate at which the source must generate its status updates to keep the monitor updated in the most timely manner. This rate was shown to be *different* from those that maximize throughput or minimize status packet delivery delay. The results of this seminal work prompted the interest of the research community, and a surge in the number of papers published on this topic can be witnessed [10, 11].

At the core of all the research works in this area, age-based metrics were used to represent the level of dissatisfaction for having aged information at the monitor. These metrics are functionals of the age process $\{\Delta(t), t \geq 0\}$. Among the widely used age-metrics, we cite the following:

- The time-average age [8]:

$$\phi_{\text{avg}}(\{\Delta(t), t \geq 0\}) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T \Delta(t) dt. \tag{2.2}$$

- The average Peak Age of Information (PAoI) [12]:

$$\phi_{\text{peak}}(\{\Delta(t), t \geq 0\}) = \lim_{K \to +\infty} \frac{1}{K} \sum_{i=1}^{K} C_i, \tag{2.3}$$

  where C_i is the i-th peak age value since $t = 0$ as shown in Fig. 2.3.

- The time-average of a non-decreasing function of age p_t [7]:

$$\phi_{\text{avg-pen}}(\{\Delta(t), t \geq 0\}) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T p_t(\Delta(t)) dt, \tag{2.4}$$

  where $p_t : [0, +\infty[\mapsto \mathbb{R}$ is a non-decreasing function. The non-decreasing property is natural and complies with the fact that fresh data are often more desired than stale/outdated data. Non-linear age functions have been shown to be closely related to numerous real-time applications where various performance measures can be cast as the time-average of a non-decreasing function of the age. Among these quantities, we cite the estimation error and auto-correlation of real-time signals [13].

Equipped with these age metrics, researchers examined the potentials of the AoI.

2.3 State of the Art and Research Trends

The benefits of the AoI metric in modeling various real-life applications are regularly brought to light, highlighting the broad scope of this metric. For example, extensive studies on the path planning problem for Unmanned Aerial Vehicles (UAV) networks have been carried in the literature based on many well-known design principles such as throughput, energy efficiency, and flight time. The authors in [14] have developed an age-based path planning for UAV networks and highlighted the

gains in information freshness compared to the traditional path planning algorithms designs. In another line of work, channel estimation problems in wireless networks were investigated [15]. In particular, in many wireless communication systems, the transmitter adapts to the instantaneous state of a time-variant wireless channel using a channel estimate at its disposal. Transmitters may use outdated channel state information due to the high cost of training. The authors in [15] were able to characterize the performance degradation due to this phenomenon based on the concept of the age of information. Recently, the potential use of the AoI to keep the citations index of researchers up to date on websites such as Google Scholar was investigated in [16]. As one can see, the span of applications that the AoI encompasses is large, and the ground is fertile for many more to come. This renders the analysis of age-based metrics in general system environments essential to have an even better understanding of the metric. This has been a fundamental focus of the research work in the area, and it is one of the main objectives that this thesis aims to achieve. Among these research works, the ones most pertinent to this thesis can be categorized into three main categories:

1. Age-optimization of queuing systems
2. Age-based scheduling
3. Remote estimation applications

2.3.1 Queuing Systems: Analysis and Optimization

As the case for the standard throughput and delay metrics, the first step to better understand a new metric and its dynamics is to analyze and optimize it in general queuing systems. Ever since its revival in [8], the AoI has been examined in a broad span of queuing systems. In [12], the authors broke out of the standard FCFS queuing discipline and have shown that the management of packets can further minimize the AoI compared to the FCFS discipline. Following the same footsteps, the Last-Generated-First-Served (LGFS) policy was proven optimal for minimizing the AoI in both single and multi-hop networks [17]. The benefits of having parallel servers on the AoI performance were explored in [18]. The effect of packet losses on the AoI was analyzed in [19]. Source coding and channel coding schemes for minimizing the age were also proposed in [20–22].

Status updates systems where the source is constrained by energy consumption were heavily studied [23–26]. For example, the authors in [23] investigated the case where the source's update rate cannot exceed a predefined limit due to battery considerations. Energy harvesting systems have also attracted a fair share of research attention. In [25], Yates examined the case where a stochastic energy harvesting system powers the source. He showed that the age-optimal policy is lazy. In other words, following a service completion, the service facility is frequently left idle even though the server may have sufficient energy to submit an update. The results disseminated by Yates in [25] were surprising as it was believed a new packet should

be sent whenever available energy levels allow it. Capitalizing on these results, Sun et al. investigated a delay channel where the source can generate packets at will [7]. The system is reported in Fig. 2.4. They were able to show similar results that a zero-wait sampling policy can be far from age-optimal. In the zero-wait sampling policy, a new sample is generated and transmitted once the previous update is delivered, and an Acknowledgement (ACK) packet is received. The results in [7] suggest that the source may need to wait before generating and transmitting the packet. The waiting time depends on the distribution of the channel delay and the current age at the monitor. Similar results were later found for the stochastic arrivals case in [27].

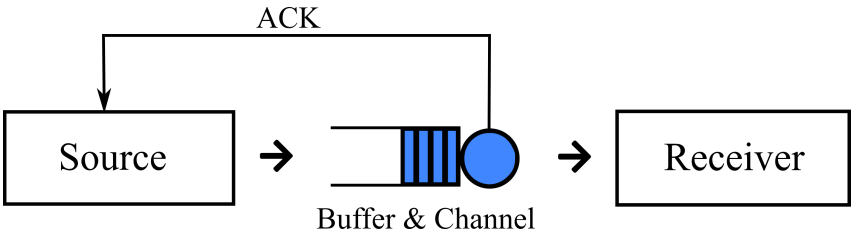


Figure 2.4: Generate at will model with ACK.

There also exist several recent studies on status updates with multiple priority classes. This is a natural consequence of the vast amount of real-life scenarios where information streams are assigned different priorities based on their sensitivity. In [28], the authors considered multiple information streams, each with a different priority, and sharing a common service facility with null or one waiting room in the queue that is shared by the streams. They studied the case where a high priority packet would preempt a lower priority packet, which is then discarded. They derived an expression of each stream's average age, and the arrival rate of each stream was then optimized accordingly. Most recently, closed forms of the average peak AoI were found in M/M/1/1 settings, where streams are assigned different priorities [29].

Due to the importance of multi-class environments, we examine in Chapter 3 of this thesis a priority-based queuing system. We analyze the system's average age and provide results on the interaction between the different classes. These results will be the backbone of our proposal of a new age-optimization framework for multi-class settings.

2.3.2 Age-Based Scheduling

In the majority of real-time applications, competing information streams share the transmission channel where the available resources can be scarce. The scarcity can be a result of battery considerations for the devices involved or physical interference limiting the number of simultaneous transmissions between information streams. As a result, an adequate assignment of the available resources to achieve the timeliness goal of the overall system has to be adopted. To that end, researchers have

widely explored a range of scheduling problems that aim to minimize age-based metrics in various resource-constrained settings. The efforts can be categorized into two main parts.

Centralized Scheduling

In *centralized* scheduling, a single dedicated scheduler handles the management and allocation of the available resources. The scheduler typically has accurate information about the state of the whole system. The available information at its disposal span 1) network topology, 2) instantaneous AoI at the monitor, 3) battery consumption budgets of each device and 4) channel conditions. Many networks fall under this category; perhaps the most notable is the standard cellular network where the base station handles the scheduling of devices within the cell.

Up to this point, researchers have examined the AoI based metrics in a variety of network settings. In [30], the authors investigated the weighted average AoI in a multi-user environment where, at most, one user can be scheduled, and channels exhibit possible decoding errors. It was shown that a greedy algorithm is optimal when users have identical channel statistics. In the asymmetric case, sub-optimal policies were proposed. Likewise, a Markov Decision Process (MDP) approximation approach was used to propose both offline and online scheduling policies in the asymmetric case [31]. In the same settings but under random arrivals, a Whittle's index policy was proposed to minimize the average AoI [32]. In [33], the authors showed that under the assumption of synchronized packets' arrival and generation times, a greedy algorithm is optimal in the multi-stream single exponential server scenario. In another line of work, the physical Signal-to-Interference-plus-Noise Ratio (SINR) model was examined in a multi-user setting. It was demonstrated that finding the minimum average age policy is NP-Hard, and a low-complexity heuristic was therefore proposed. In the case of delay channels [34], the zero-wait sampling policy was proved to minimize the total average peak AoI in multi-user settings.

To this date, new research works on age-based scheduling in centralized environments are being published. We encourage the readers to refer to [35] for a list of the most recent publications on the subject. As many systems are distributed by nature, and centralized schemes cannot be therefore adopted, distributed scheduling algorithms were investigated in the literature.

Distributed Scheduling

In many real-life scenarios, especially IoT applications, having a central entity that manages the available resources is not necessarily feasible. This infeasibility is caused by various reasons such as cost, complexity, power considerations, and the autonomous nature of devices. Therefore, instead of a centralized approach, a distributed one must be adopted where devices only use local information at their disposal to access the available resources. With the AoI being of a broad interest in

this type of application, considerable research efforts have been put on optimizing the AoI in decentralized scheduling environments.

Among these efforts, the authors in [36] studied a distributed scheduling scheme where devices that always have fresh information at their disposal access the channel with a certain probability. The access probability of each link is optimized to minimize the total average age of the network. In [37], it was shown that a distributed Round Robin (RR) multiple access scheme achieves the minimum average age when the number of devices tends to infinity for noise-free channels. In the case of noisy channels, the authors in [38] proposed a decentralized access scheme based on the Whittle's index. It was shown numerically to possess good performance in the many devices regime. In another line of work, the Carrier Sense Multiple Access (CSMA) medium access scheme was investigated in the framework of vehicular networks in [39]. In this work, the authors studied the average AoI of a vehicular network environment *numerically*. It was shown that the information age is minimized at an optimal operating point between the two extremes of maximum throughput and minimum delay.

Although CSMA is considered one of the most renowned and widely spread distributed scheduling schemes, theoretical characterization of its age optimal operating point remains an open question. To address this, we theoretically study in Chapter 4 a general CSMA environment and provide the first theoretical results on its operating point that minimizes the average AoI.

2.3.3 Remote Estimation Applications

Among the various applications of the AoI that researchers investigated, perhaps the ones that stands out most are remote estimation problems. The main reason for that is the broad applicability of these problems in real-life scenarios. In remote estimation applications, the source observes a process X_t that changes over time. X_t can be, for example, the temperature of a room, the velocity of a vehicle, or the commands of a controller. Using the status updates sent by the source, the monitor constructs an estimate of X_t, denoted by $\hat{X}_t$. The goal is to make $\hat{X}_t$ as close as possible to X_t. The standard approach to achieve this objective is to minimize the estimation error $\Pr(X_t \neq \hat{X}_t)$ or the Mean Squared Error (MSE) $\mathbb{E}[|X_t - \hat{X}_t|^2]$. Researchers have shown, in many works, that the AoI is closely related to the two objectives mentioned above. For instance, Sun et al. investigated the Minimum Mean Squared Error (MMSE) policy in a delay channel where X_t is a Wiener process [40]. They have shown that finding the MMSE sampling policy is equivalent to finding the sampling policy that minimizes the average of a non-linear function of the age. Their results hold for the case where sampling instants are independent of the values of X_t. Similarly, in Network Controlled Systems (NCS) where the plant and the controller are governed by a Linear Time-Invariant (LTI) system, the minimization of a particular non-linear function of the AoI leads to the minimization of the MSE [41]. In another line of work, maximizing the mutual information between the two processes X_t and $\hat{X}_t$ when X_t is a stationary Markovian process

was also shown to be related to the minimization of the average of a non-linear age function [42].

At the same time, shortcomings of the AoI were also put into perspective for particular real-time estimation problems. For instance, when X_t is a Wiener process and sampling times are allowed to depend on X_t, the MMSE policy in a delay channel is not equivalent to a minimization of an age-based metric [40]. Similarly, the AoI was shown in [43] to be sub-optimal in minimizing the estimation error over channels with potential decoding errors when X_t is a Markov process. The main explanation to these shortcomings is that the AoI, by definition, does not depend on X_t and $\hat{X}_t$. These observations prompted efforts to propose new performance metrics that deal with the shortcomings of the AoI. Among these efforts, a time-based metric dubbed as the Age of Synchronization (AoS) was introduced in the framework of content caching [44, 45]. Specifically, the AoS is zero when the transmitter has no packets to send. It grows linearly with time when the transmitter side generates a new packet. Although the AoS includes the packets generation as a factor, it does not depend on X_t and $\hat{X}_t$, which may limit its usage in remote estimation applications. In another work [46], the authors proposed different *effective* age metrics for which a lower effective age should undoubtedly lead to a lower prediction error. For example, the notion of *sampling age* was introduced, which was defined as the age relative to an ideal sampling pattern $g(t)$ that minimizes the error. However, finding the optimal pattern $g(t)$ was deemed to be far from being trivial.

In Chapter 5 of this thesis, on top of the AoI's shortcomings, we shed light on the shortcomings of the conventional error penalty functions (e.g., estimation error, mean squared error). Accordingly, we propose a new performance metric that deals with these shortcomings, and we showcase its advantages in many real-life applications.

2.4 Thesis Outline and Contributions

This thesis is divided into two parts. In Part I, we focus on optimizing age-based metrics in various communication systems and network settings. Specifically:

In Chapter 3 (Multi-Class Multi-Stream Scheduling), we consider a transmission scheduling problem, in which several streams of status update packets with diverse priority levels are sent through a shared channel to their destinations. We find a closed-form expression of each stream's average age for Poisson arrivals and exponential service times. Using these expressions, we provide insights on the interaction between the different classes. These insights suggest the need for a new age-optimization framework in multi-class systems. To that end, 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 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 pol-

icy 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.

In Chapter 4 (Average Age Minimization in a CSMA Environment), we investigate a network where N links contend for the channel using the well-known CSMA scheme. CSMA is a class of simple and distributed multiple access protocol that is seen as one of the most popular distributed Medium-Access-Control (MAC) schemes in wireless networks (e.g., CSMA is the basic medium access algorithm in IEEE 802.11 [47]). In this class of schemes, a transmitter waits for a certain duration of time before transmitting, called the back-off time. In this transmission environment, we derive closed-form expressions of the average AoI in distinct packet arrivals settings. Equipped with these expressions, we formulate the problem of minimizing the average age by calibrating the back-off time of each link. By analyzing its structure, we convert the formulated optimization problem to an equivalent convex problem that we derive its optimal solution. Insights on the interaction between links and numerical implementations of the optimized CSMA scheme in an IEEE 802.11 environment are also presented. Furthermore, to improve the performance of the optimized CSMA scheme, we propose a modified version of the scheme by giving each link the freedom to transition to SLEEP mode. The proposed approach provides a way to reduce the burden on the channel when possible. This leads to an improvement in the performance of the network.

In Part II of the thesis, we propose a new performance measure that addresses the shortcomings of the AoI and the conventional, and widely employed, error penalty functions. In particular:

In Chapter 5 (Age of Incorrect Information: Analysis and Optimization), we introduce a new performance metric in the framework of status updates systems which we refer to as the Age of Incorrect Information (AoII). This new metric deals with the shortcomings of both the AoI and the error penalty functions as it neatly extends the notion of fresh updates to that of fresh "informative" updates. The word informative in this context refers to updates that bring new and correct information to the monitor side. The AoII also captures the deteriorating effect wrong information can have with time on the system. After motivating the new metric, we find the optimal sampling policies that minimizes the AoII in both unconstrained and resource-constraint scenarios. These policies are then implemented and their advantages compared to both the age-optimal and error-optimal policies in a variety of real-life applications are highlighted.

Finally, we present our conclusions and future works that can be undertaken as a continuation of this thesis. We note that each chapter above contains its own mathematical notation.

2.5 Publications

The following publications were produced during the course of this thesis [48–63]. The results/details which are either fully or partially provided in this manuscript are marked with *.

Journal Articles

[48]* 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, pp. 1–14, 2020.

[49]* A. Maatouk, M. Assaad, and A. Ephremides, “On The Age of Information in a CSMA Environment," IEEE/ACM Transactions on Networking, vol. 28, no. 2, pp. 818–831, 2020.

[50] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “On The Optimality of The Whittle’s Index Policy For Minimizing The Age of Information," revisions stages in IEEE Transactions on Wireless Communications.

[51] A. Maatouk, S. E. Hajri, M. Assaad, and H. Sari, “On Optimal Scheduling for Joint Spatial Division and Multiplexing Approach in FDD Massive MIMO," IEEE Transactions on Signal Processing, vol. 67, no. 4, pp. 1006–1021, 2019.

[52] A. Maatouk, M. Assaad, and A. Ephremides, “Energy Efficient and Throughput Optimal CSMA Scheme," IEEE/ACM Transactions on Networking, vol. 27, no. 1, pp. 316–329, 2019.

[53] A. Maatouk, E. Caliskan, M. Assaad, M. Koca, G. Gui, and H. Sari, “Frequency-Domain NOMA With Two Sets of Orthogonal Signal Waveforms," IEEE Communications Letters, vol. 22, no. 5, pp. 906–909, 2018.

Conference Papers

[54]* 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), pp. 1–8, 2020.

[55]* A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides “Asymptotically Optimal Scheduling Policy For Minimizing The Age of Information," in 2020 IEEE International Symposium on Information Theory (ISIT), pp. 1747–1752, 2020.

[56]* A. Maatouk, M. Assaad and A. Ephremides, “Minimizing The Age of Infor-

mation in a CSMA Environment," in 2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT), pp. 1– 8, 2019.

[57] 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), pp. 102–108, 2019.

[58]* 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), pp. 325–329, July 2019.

[59] A. Maatouk, M. Assaad and A. Ephremides, “The Age of Updates in a Simple Relay Network," in 2018 IEEE Information Theory Workshop (ITW), pp. 1–5, 2018.

[60] E. Caliskan, A. Maatouk, M. Koca, M. Assaad, G. Gui and H. Sari, “A Simple NOMA Scheme with Optimum Detection," in 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2018.

[61] J. Denis, A. Maatouk, S. E. Hajri and M. Assaad, “Stay Longer at the Network’s Edge: A Novel Proactive Caching Policy through Sojourn Time," in 2018 IEEE Global Communications Conference (GLOBECOM), pp. 1–6, 2018.

[62] A. Maatouk, S. E. Hajri, M. Assaad, H. Sari and S. Sezginer, “Graph Theory Based Approach to Users Grouping and Downlink Scheduling in FDD Massive MIMO," in 2018 IEEE International Conference on Communications (ICC), pp. 1–7, 2018.

[63] H. Sari, A. Maatouk, E. Caliskan, M. Assaad, M. Koca and G. Gui, “On the foundation of NOMA and its application to 5G cellular networks," in 2018 IEEE Wireless Communications and Networking Conference (WCNC), pp. 1– 6, 2018.

Part I

Age of Information Optimization

3 | Multi-Class Multi-Stream Scheduling

3.1 Overview

This chapter investigates a multi-class system where N streams of I priority levels share a common service facility. As explained in the introduction, priority-based queuing systems are abundant in real-life scenarios (e.g., vehicular networks, smart homes, etc.). Accordingly, these systems have recently gained significant research attention within the AoI community [28, 29]. The majority of the efforts lay mainly in finding closed-form expressions of the average AoI/PAoI in a particular scenario and for a specific arrival/transmission model. Accordingly, the question of what is the age-optimal scheduling policy in a multi-class priority-based scheduling scenario remains open. In this chapter, we find an answer to this question. To that end, we summarize in the following the key contributions of this chapter:

- We analyze the average AoI when packets' inter-arrival times and service times are exponentially distributed. Unlike the work in [28], we consider that each stream has its own buffer space. The investigation of this setting will highlight the interactions between the different classes and let us pave the way for proposing a new framework for AoI-optimality in these multi-class settings.
- We introduce the notion of Lexicographic optimality for the age minimization framework, which we will refer to as the lex-age-optimality. The lex-age-optimality elegantly captures both the age-optimality and the order of time-cruciality between the streams in a general multi-class scheduling scenario. This approach guarantees that the low priority streams' performance is optimized while ensuring that the facility grants high priority streams the best possible service.
- In the case of a single server with i.i.d. exponential service times, we propose the PP-MAF-LGFS scheduling policy. Using a sample-path argument, we show that this policy is lex-age-optimal. Our lex-age-optimality results are not constrained to the traditional minimization of the average AoI and PAoI

frameworks previously adopted in [28, 29]. 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. Our lex-age-optimality results are also not bound to any traffic arrival distribution. Moreover, they hold when the priority classes have distinct traffic patterns and different dissatisfaction levels of the aged information. This showcases our results' broad scope as classes typically represent diverse applications, each with its data timeliness requirements. For example, we could be interested in minimizing the average AoI for a class and the average age penalty for another.

The rest of the chapter is organized as follows: Section 3.2 describes the system model and the mathematical tools used in the chapter. In Section 3.3, we derive a closed-form expression of each stream's average age for Poisson arrivals and exponential service times. In Section 3.4, we introduce the notion of lex-age-optimality and propose a lex-age-optimal policy in the single exponential server settings. Numerical implementations are also presented in both Section 3.3 and 3.4 while Section 3.5 concludes the chapter.

3.2 General Setup and Mathematical Tools

3.2.1 Preliminaries on Stochastic Hybrid Systems

Throughout the AoI literature, two main tools can be found for analyzing the AoI for a particular system:

1. The graphical area decomposition method
2. The Stochastic Hybrid System (SHS) method

These methods are used to derive a closed-form expression of the average age under a specific scheduling policy. Afterward, the system's parameters are optimized to minimize the average AoI. The graphical area decomposition method has been first introduced by Kaul et al. in [8]. Since then, various work has adopted the method for numerous other system settings (e.g., [12, 18]). To understand this approach, we recall from eq. (2.2) that the average age of a certain system is nothing but the area below the sawtooth curve depicting the evolution of the instantaneous system's age $\Delta(t)$. To that end, let us consider a random time instant τ. By noting the sawtooth trend of the AoI, and as seen in Fig. 3.1, the area below the curve up to the time instant τ can be divided into $N(\tau)+1$ parts: 1) the polygon Q_1, 2) the $N(\tau)-1$ trapezoids, and 3) the triangle $\tilde{Q}_{N(\tau)}$. $N(\tau)$ denotes the number of drops in the age that takes place up to the time instant τ. Next, these areas are written in function of the system's dynamics, such as packet arrivals, queuing delays, and transmission times. By letting τ tend to $+\infty$, and using these areas' expressions, the average age can then be deduced. Depending on the system's in question, this approach can

allow us to find a closed-form expression of the average age. However, depending on the complexity of the system, this method can hit a roadblock. Accordingly, several other tools have been proposed for finding these closed-form expressions, the most notable of them being the SHS tool [64–67].

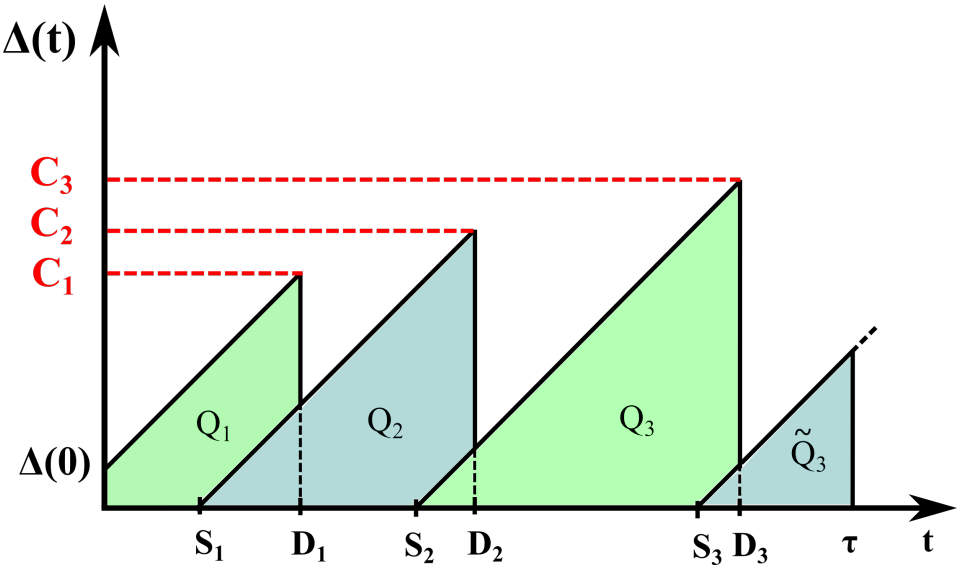


Figure 3.1: Illustration of the graphical area decomposition method.

Stochastic hybrid systems involve the coupling of discrete, continuous, and probabilistic phenomena. Because of their versatility and generality, methods for modeling and analysis of stochastic hybrid systems have proved invaluable in a wide range of applications, including finance, air traffic control, biology, telecommunications, and embedded systems [68]. The SHS tool was introduced to the AoI literature by Yates et al. in [64] as a means to circumvent difficulties involved in the AoI analysis using the graphical area decomposition approach. In particular, the graphical approach for finding the average AoI of the system can be challenging in lossy systems where packets might be preempted and discarded [69]. The SHS approach involves modeling the system in question through the states $(q(t), \boldsymbol{x}(t))$ where:

- $q(t)$ is a process of discrete nature that captures the various events that can take place in the system (e.g., a packet arrival, a successful transmission, etc.). We denote by $\mathbb{Q}$ the discrete set of values that $q(t)$ can have.
- $\boldsymbol{x}(t) \in \mathbb{R}^N$ is a process of continuous nature that represents the evolution of the age in the network. The dimension N will depend on the system in question, as will be seen in applying this tool in several systems in our thesis.

Although the SHS theory is rich and may include various possibilities for $q(t)$, we focus in this thesis on the case where $q(t)$ is a Markov process. In other words, $q(t)$ can be represented graphically by a Markov chain $(\mathbb{Q}, \mathbb{L})$. To that end, each

state $q \in \mathbb{Q}$ is a vertex in the chain and each transition $l \in \mathbb{L}$ is a directed edge (q_l, q'_l) with a transition rate $\lambda^{(l)}\delta_{q_l,q(t)}$. The Kronecker delta function assures that this transition l can only occur when the discrete process $q(t)$ is equal to q_l. For each state q, we define the incoming and outgoing transitions set respectively as:

$$\mathbb{L}'_q = \{l \in \mathbb{L} : q'_l = q\}, \quad \mathbb{L}_q = \{l \in \mathbb{L} : q_l = q\}. \tag{3.1}$$

The interest in SHS stems from the fact that the discrete process transitions will induce a reset to the continuous process. More specifically, when a transition l takes place, the discrete process jumps to another state q'_l and a discontinuous jump in the continuous process $\boldsymbol{x'} = \boldsymbol{x}\boldsymbol{A}_l$ is seen. We recall from Section 2.2, that this is exactly the behavior of the AoI curve, where drops happen upon delivery of a packet to the monitor. The matrix $\boldsymbol{A}_l \in \mathbb{R}^N \times \mathbb{R}^N$ is referred to as the transition l reset maps. These reset maps will be pivotal to the proper modeling of the evolution of the age process. For example, the matrix $\boldsymbol{A}_l$ will allow us to model a drop in the age at the monitor when a packet is successfully delivered.

Finally, to fully capture the evolution of the age process in each state $q \in \mathbb{Q}$, we recall that the age can only increase linearly with time. To incorporate this in the analysis, we let the continuous process $\boldsymbol{x}(t)$ verify in each state $q \in \mathbb{Q}$ the following first-order differential equation: $\dot{\boldsymbol{x}}(t) = \boldsymbol{b}_q$. In this context, $\boldsymbol{b}_q$ is a binary vector whose i-th component b^i_q is equal to 1 if the age $x_i(t)$ increases at a unit rate when the system is in state q (i.e., $\dot{x}_i(t) = 1$) and is equal to 0 if the age keeps the same value in this state ($\dot{x}_i(t) = 0$). To calculate the average age of the system through SHS, the following quantities for each state $q \in \mathbb{Q}$ need to be defined:

$$\pi_q(t) = \mathbb{E}[\delta_{q,q(t)}] = P(q(t) = q), \tag{3.2}$$

$$\boldsymbol{v}_q(t) = \mathbb{E}[\boldsymbol{x}(t)\delta_{q,q(t)}], \tag{3.3}$$

where $\pi_q(t)$ is the Markov chain's state probability and $\boldsymbol{v}_q(t) \in \mathbb{R}^N$ denotes the correlation between the age process and the discrete state of the system q. To ensure the existence and uniqueness of a steady-state distribution of the Markov chain, we assume that the Markov chain $q(t)$ is ergodic. To that end, we define the steady-state probability vector $\overline{\boldsymbol{\pi}} = [\overline{\pi}_q]_{q\in\mathbb{Q}}$ as the solution to the following general balance equations:

$$\overline{\pi}_q\Big(\sum_{l\in\mathbb{L}_q} \lambda^{(l)}\Big) = \sum_{l\in\mathbb{L}'_q} \lambda^{(l)}\overline{\pi}_{q_l}, \quad \forall q \in \mathbb{Q}, \tag{3.4}$$

$$\sum_{q\in\mathbb{Q}} \overline{\pi}_q = 1. \tag{3.5}$$

As it has been shown in [64], the correlation vector $\boldsymbol{v}_q(t)$ converges in this ergodic case to a limit $\overline{\boldsymbol{v}}_q$ such that:

$$\overline{\boldsymbol{v}}_q\Big(\sum_{l\in\mathbb{L}_q} \lambda^{(l)}\Big) = \boldsymbol{b}_q\overline{\pi}_q + \sum_{l\in\mathbb{L}'_q} \lambda^{(l)}\overline{\boldsymbol{v}}_{q_l}\boldsymbol{A}_l, \quad \forall q \in \mathbb{Q}. \tag{3.6}$$

Building on this, we can deduce that

$$\mathbb{E}[\boldsymbol{x}] = \lim_{t\to+\infty} \mathbb{E}[\boldsymbol{x}(t)] = \lim_{t\to+\infty} \sum_{q\in\mathbb{Q}} \mathbb{E}[\boldsymbol{x}(t)\delta_{q,q(t)}] = \sum_{q\in\mathbb{Q}} \overline{\boldsymbol{v}}_q. \tag{3.7}$$

Based on the aforementioned results from [64], we present the following theorem that summarizes all that have been previously stated.

Theorem 3.1. *When the Markov chain $q(t)$ is ergodic and admits $\overline{\boldsymbol{\pi}}$ as stationary distribution, if we can find a non-negative matrix $\overline{\boldsymbol{v}} = [\overline{\boldsymbol{v}}_q]_{q\in\mathbb{Q}}$ such that $\overline{\boldsymbol{v}}_q$, $\forall q \in \mathbb{Q}$ is the solution of eq. (3.6), then the SHS is stable and the average age of the component i of $\boldsymbol{x}$ is:*

$$\overline{\Delta}_i = \sum_{q\in\mathbb{Q}} \overline{v}_{qi}. \tag{3.8}$$

The above theorem's results will be utilized to derive closed-form expressions of the average age in this chapter and several other scenarios throughout the thesis.

3.2.2 Notations and Definitions

Throughout this chapter, 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 our subsequent age-optimality analysis.

Definition 3.1. *Stochastic Ordering of Random Variables [70]: A random variable X is said to be stochastically smaller than a random variable Y, denoted by $X \leq_{st} Y$, if* $\Pr(X > t) \leq \Pr(Y > t), \ \forall t \in \mathbb{R}$.

Definition 3.2. *Upper Sets: A set $\mathcal{U} \subseteq \mathbb{R}^N$ is called upper if $\boldsymbol{y} \in \mathcal{U}$ whenever $\boldsymbol{x} \leq \boldsymbol{y}$ and $\boldsymbol{x} \in \mathcal{U}$.*

Definition 3.3. *Stochastic Ordering of Random Vectors [70]: 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*

$$\Pr(\boldsymbol{X} \in \mathcal{U}) \leq \Pr(\boldsymbol{Y} \in \mathcal{U}), \quad \textit{for all upper } \mathcal{U} \subseteq \mathbb{R}^N. \tag{3.9}$$

Definition 3.4. *Stochastic Ordering of Stochastic Processes [70]: A stochastic process* $\{X(t), t \geq 0\}$ *is said to be stochastically smaller than a 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* $t_1 < t_2 < \ldots < t_m \in \mathbb{R}^+$

$$(X(t_1), X(t_2), \ldots, X(t_m)) \leq_{st} (Y(t_1), Y(t_2), \ldots, Y(t_m)). \tag{3.10}$$

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

$$\mathbb{V} = \{g : [0, \infty) \mapsto \mathbb{R} \text{ is Lebesgue measurable}\}. \tag{3.11}$$

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, [70]

$$\mathbb{E}[\phi(\{X(t), t \geq 0\})] \leq \mathbb{E}[\phi(\{Y(t), t \geq 0\})] \tag{3.12}$$

holds for every non-decreasing functional ϕ for which the expectations in (3.12) exist.

3.2.3 System Model

Queuing Model

Consider the status-update system illustrated in Fig. 3.2, where N streams of update packets are sent through a common service facility. Each update stream has a buffer space, which can be infinite or finite. The server can process at most one packet at a time. The packet service times are i.i.d. across streams and time. The information streams are divided into I priority classes, with streams of the same class i having the same priority. Each information stream is indexed by two components (i, j), where i denotes the class index and j denotes the stream index within class i. The classes are indexed in 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 can have different source and/or destination nodes.

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 $S_n^{i,j} \leq A_n^{i,j} \leq D_n^{i,j}$. Let π represents a scheduling policy that determines the packets being sent over time. Let Π denotes the set of all *causal* scheduling policies, i.e., where the decisions are taken without any knowledge of the future. 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.

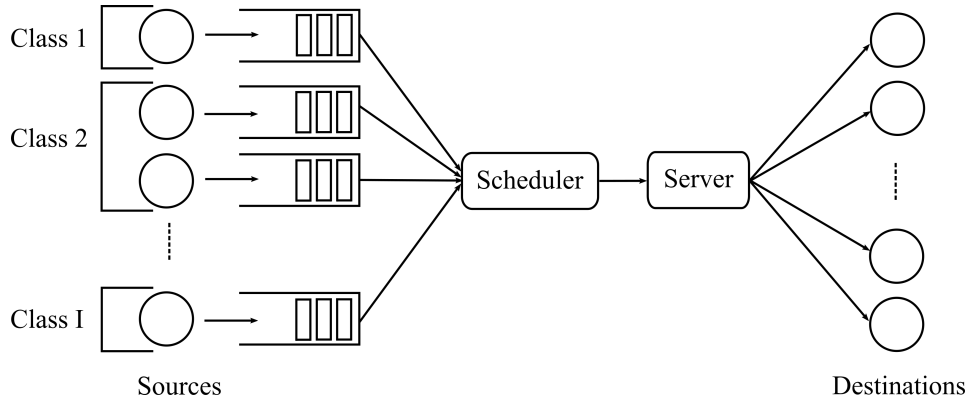


Figure 3.2: Multi-class system model.

Age Penalty Functions and Functionals

We denote the instantaneous age of information of stream (i, j) at time instant t as:

$$\Delta^{i,j}(t) = t - \max\{S_n^{i,j} : D_n^{i,j} \leq t, \;\; n = 1, 2, \ldots\}, \tag{3.13}$$

We let $\mathbf{\Delta}^i(t) = (\Delta^{i,1}(t), \ldots, \Delta^{i,J_i}(t))$ denote the age vector at time t of all streams belonging to class i. Additionally, we let $\mathbf{\Delta}(t) = (\mathbf{\Delta}^1(t), \ldots, \mathbf{\Delta}^I(t))$ denote the age vector of all streams at time t.

As it has been previously discussed in Section 2.2, age-based metrics were used throughout the literature to represent the level of dissatisfaction for having aged information at the monitor. In this chapter, we calibrate this approach to the multi-class system. In particular, we will use both age function and funtionals to represent the dissatisfaction with the staleness of data on a per-class basis. To that end, we introduce an age penalty function $p_t \circ \mathbf{\Delta}^i(t)$ that represents the level of dissatisfaction with the aged information at time t for class i, where $p_t : \mathbb{R}^{J_i} \mapsto \mathbb{R}$ is a non-decreasing function of $\mathbf{\Delta}^i(t)$. Some commonly used age penalty functions that can be adopted are listed below.

- The sum age of the class i streams:

$$p_{\text{sum}} \circ \mathbf{\Delta}^i(t) = \sum_{j=1}^{J_i} \Delta^{i,j}(t). \tag{3.14}$$

- The maximum age of the class i streams:

$$p_{\max} \circ \mathbf{\Delta}^i(t) = \max_{j=1,\ldots,J_i} \Delta^{i,j}(t). \tag{3.15}$$

- The average age threshold violation of the J_i streams:

$$p_{\text{exceed}-\alpha} \circ \mathbf{\Delta}^i(t) = \frac{1}{J_i} \sum_{j=1}^{J_i} \mathbb{1}_{\{\Delta^{i,j}(t) > \alpha\}}. \tag{3.16}$$

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:

$$p_{\text{pen}} \circ \mathbf{\Delta}^i(t) = \sum_{j=1}^{J_i} g(\Delta^{i,j}(t)), \tag{3.17}$$

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 [7].

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

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

This class of penalty functions $\mathcal{P}_{\text{sym}}$ is fairly large, and include the provided age penalty functions (3.14)-(3.17). 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 our considered penalty functions.

In addition to age penalty functions, we use non-decreasing age penalty 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. An example of these functionals is the time-average age penalty, previously reported in eq. (2.2):

$$\phi_{\text{avg}}(\{p_t \circ \mathbf{\Delta}^i(t), t \geq 0\}) = \lim_{T \to +\infty} \frac{1}{T} \int_0^T p_t \circ \mathbf{\Delta}^i(t) dt. \tag{3.18}$$

We consider in this chapter that different priority classes can have distinct age penalty functions and functionals. This is of paramount importance as each priority class typically represents a different application, each with its own data timeliness requirements. For example, in a vehicular network, time-crucial safety data related to vehicle position should be delivered promptly. Typically, the system performance is affected by the maximum age of the delivered updates. Accordingly, we can choose the maximum age penalty function $p_{\max}$ and the average age penalty function ϕ_{avg} for this class of traffic. On the other hand, updates on gas tank levels require an average timely delivery. Consequently, we can choose the penalty function p_{sum} and the time-average age penalty functional ϕ_{avg} for this class.

In the sequel, we use $\{\mathbf{\Delta}^i_\pi(t), t \geq 0\}$ and $\{p_t \circ \mathbf{\Delta}^i_\pi(t), t \geq 0\}$ to represent the stochastic age process and 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 $\pi \in \Pi$.

3.3 AoI Analysis in Multi-class Environments

In this first part of the chapter, we focus on a particular multi-class environment under a fixed scheduling policy. The goal is to shed light on the interaction between the various classes to pave the way for the proposal of a new age-optimization framework in multi-class systems. To that end, let us consider the case where $J_i = 1$ for all classes $i = 1, \ldots, I$. In other words, all streams have distinct priorities ($N = I$). Consequently, we drop the index j of each stream's index throughout this section. We suppose that the generation time and arrival time of all packets in the network are identical. In particular,

$$S_n^i = A_n^i, \quad \text{for } i = 1, \ldots, I, \quad \text{and } n = 1, 2, \ldots. \tag{3.19}$$

Put in another way, there is no delay between the generation of the packet and its arrival to the queue. We consider that each of these packets' transmission time is exponentially distributed with rate μ. We also assume that the packets arrival of each stream is a Poisson process with an average rate λ.

In this multi-class scenario, we investigate a specific scheduling policy. In this policy, a higher priority packet will always preempt the service of a lower priority stream. The preempted packet is then stored in its own single buffer space. Any new arrival for a specific class will replace any packet of the same class already in the system. This includes the case where the packet is being served. This is motivated by the fact that a preemptive M/M/1/1 scenario was shown to minimize the average age in the case of exponential transmission times [71].

We are interested in analyzing the average age of each stream to provide insights on the interaction between the classes. In other words, we utilize the sum penalty function $p_t = p_{\text{sum}}$, and the time-average age penalty functional $\phi = \phi_{\text{avg}}$. In the sequel, the average age of stream i (or equivalently, class i) for $i = 1, \ldots, I$ will be denoted by $\overline{\Delta}_i^{\text{WQ}}$. WQ refers to "With Queues" to signal that there is an available buffer space (waiting room) for each of the individual streams.

3.3.1 Theoretical and Numerical Analysis

Average Age Calculation

In this section, we leverage the SHS approach to calculate the average age of each stream in the system. To make calculations easier, we forgo studying the age process of all streams simultaneously. Instead, we examine the perspective of stream i for $i = 1, \ldots, I$. We recall that the classes are indexed in decreasing order of priority. With that in mind, and by considering the adopted policy, we can assert that stream i can be preempted while in service by $i - 1$ higher priority streams. Based on this, we can define the discrete states $\mathbb{Q} = \{0, 1, 2, \ldots, i - 1\}$ where:

- $q(t) = 0$ if the server is serving the stream of interest.

- $q(t) = j$, with $1 \leq j \leq i-1$ if there are j packets of higher priority streams that have to be served before the server can work on the stream of interest.

The continuous-time process is defined as $\boldsymbol{x}(t) = [x_0(t), x_1(t)]$, where $x_0(t)$ is the age of the stream of interest i at time t, and $x_1(t)$ is the age of the packet that is being served (or waiting in the buffer to be served upon service completion of higher priority streams) of stream i at time t. To simplify the average age calculations, we suppose that if there are no packets of stream i in the system, a "fake" update packet is considered to be available. A fake update packet is defined as a packet with the same time-stamp as the monitor's previously received packet. Introducing the fake update will not change the average age calculation. However, it will provide mathematical benefits as it allows the reduction of the state space of $\mathbb{Q}$ since the availability of a packet of stream i in the system does not need to be monitored. Our goal is to apply Theorem 3.1 to find the vectors $\overline{\boldsymbol{v}}_q = [\overline{v}_{q0}, \overline{v}_{q1}]$, $\forall q \in \mathbb{Q}$ that will enable us to find the average age of the stream i. In order to do this, we present in the following table the transitions between the discrete states and the reset maps they induce on the age process $\boldsymbol{x}(t)$.

l	$q_l \to q'_l$	$\lambda^{(l)}$	$\boldsymbol{x}\boldsymbol{A_l}$	$\boldsymbol{A_l}$	$\boldsymbol{v}_{q_l}\boldsymbol{A_l}$
1	$0 \to 0$	λ	$[x_0, 0]$	$\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$	$[v_{00}, 0]$
2	$1 \to 1$	λ	$[x_0, 0]$	$\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$	$[v_{10}, 0]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	$i-1 \to i-1$	λ	$[x_0, 0]$	$\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$	$[v_{(i-1)0}, 0]$
$i+1$	$0 \to 1$	$(i-1)\lambda$	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[v_{00}, v_{01}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2i-1$	$i-2 \to i-1$	λ	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[v_{(i-2)0}, v_{(i-2)1}]$
$2i$	$i-1 \to i-2$	μ	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[v_{(i-1)0}, v_{(i-1)1}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$3i-2$	$1 \to 0$	μ	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[v_{10}, v_{11}]$
$3i-1$	$0 \to 0$	μ	$[x_1, x_1]$	$\left(\begin{smallmatrix}0&0\\1&1\end{smallmatrix}\right)$	$[v_{01}, v_{01}]$

Table 3.1: Stochastic hybrid system description.

We provide in the following a detailed explanation of each transition reported in Table 3.1:

1. The set of transitions from $l = 1$ till $l = i$ represents a new packet arrival for stream i. As detailed in our description of the adopted policy, a new packet will replace the packet of stream i in the waiting room. In the case where the server is already serving a packet of the stream of interest, the newly arrived

packet will preempt its service and take its place. Therefore, we can see that this transition will not affect the age of this stream at the monitor x_0. However, the age of the system's packet x_1 falls to 0.

2. The transitions set spanning from $l = i + 1$ till $l = 2i - 1$ corresponds to packets arrivals from higher-priority streams. In state $q = 0$, a packet of the stream of interest (either real or fake) is being served. As there are $i-1$ higher priority streams, any arrival from *either* of them will preempt the service of this packet. The preempted packet is brought back to its waiting room. This transition has a rate of $(i-1)\lambda$. Now, in state $q = 1$, there is already one higher priority packet being served. Let us suppose that it belongs to a stream $k < i$. We do not care about its exact priority as we only care about it being of a higher priority than the stream of interest. A new arrival for stream k will replace the packet currently in service. However, it will not affect the system from the point of view of stream i as there was already a packet for stream k being served. Therefore, this transition is omitted. However, a new packet arrival for the $i-2$ remaining higher priority streams will affect the system as there would be 2 packets to be served ahead of the packet of stream i. This transition has a rate of $(i-2)\lambda$ and will take the system from $q = 1$ to $q' = 2$. The same reasoning goes on until the last transition $l = 2i - 1$. All these transitions will not affect the stream of interest's age process and, therefore, $\boldsymbol{A}_l = \boldsymbol{I}_2$.

3. The transitions spanning from $2i$ till $3i-2$ corresponds to the server finishing the transmission of a higher-priority stream packet. These transitions have a rate of μ and will not affect the age process of the stream of interest (i.e., $\boldsymbol{A}_l = \boldsymbol{I}_2$). The last transition $l = 3i - 1$ takes place when a packet of the stream of interest finishes being served. This will reset the age at the monitor to that of the delivered packet $x'_0 = x_1$. A fake update is then generated with the same age as the previously transmitted one $x'_1 = x_1$.

As for the differential equations that portray the evolution of the age process in each discrete state, we have that in each state $q \in \mathbb{Q}$, $x_0(t)$ and $x_1(t)$ increase at a unit rate:

$$\boldsymbol{b}_q = [1 \;\; 1], \quad \forall q \in \mathbb{Q}. \tag{3.20}$$

To apply Theorem 3.1, we start by investigating the stationary distribution of the Markov Chain that models the transitions reported in Table 3.1. To do so, we provide the following proposition.

Proposition 3.1. *The continuous time Markov chain is irreducible, time-reversible and admits* $\overline{\pi}(k;\lambda,\mu)$ *as stationary distribution for any state* $0 \leq k \leq i-1$ *where:*

$$\overline{\pi}_k = \frac{\lambda^k (i-1)!}{\mu^k (i-1-k)!}\overline{\pi}_0, \tag{3.21}$$

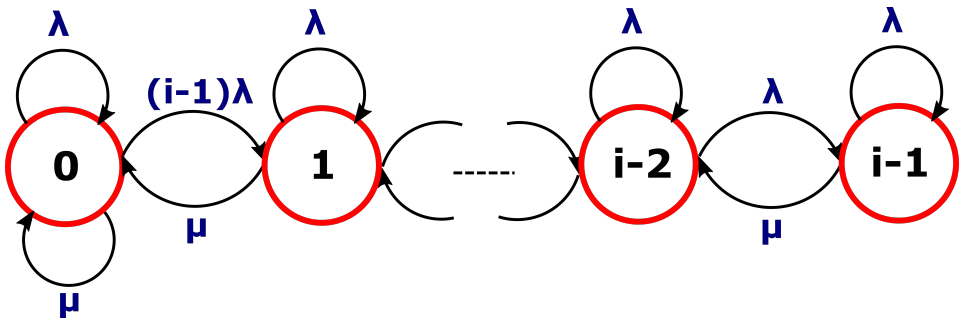


Figure 3.3: Illustration of the stochastic hybrid systems Markov chain.

and:

$$\overline{\pi}_0 = \frac{1}{\sum_{k=0}^{i-1} \frac{\lambda^k (i-1)!}{\mu^k (i-1-k)!}}. \tag{3.22}$$

Proof. We proceed with proving the proposition by induction:

- For $k = 1$, we have from eq. (3.21) that $\overline{\pi}_1 = \frac{\lambda(i-1)}{\mu}\overline{\pi}_0$. By formulating the general balance equation at state $k = 0$, we have that $\overline{\pi}_0(\lambda(i-1)) = \mu\overline{\pi}_1$ and the proposition is therefore true for $k = 1$.
- We suppose that the proposition is valid up to $k \leq i-2$ and we formulate the general balance equation at state k:

$$\overline{\pi}_k(\mu + \lambda(i-1-k)) = \lambda(i-k)\overline{\pi}_{k-1} + \mu\overline{\pi}_{k+1}. \tag{3.23}$$

  By substituting $\overline{\pi}_k$ and $\overline{\pi}_{k-1}$ by their supposed values, we can verify that $\overline{\pi}_{k+1} = \frac{\lambda^{k+1}(i-1)!}{\mu^{k+1}(i-k-2))!}\overline{\pi}_0$ which concludes our proof. The time-reversibility can be easily verified by showing that the stationary distribution provided in eq. (3.21) satisfies the detailed balance equations.

□

Equipped with this proposition, we can proceed with calculating the average age of stream i. On this note, we provide the following theorem.

Theorem 3.2. *The average age of stream i is*[1]*:*

$$\overline{\Delta}_i^{\mathrm{WQ}} = \overline{v}_{00} + \sum_{k=1}^{i-1} \overline{v}_{k0}, \tag{3.24}$$

[1]One can notice that when $i = 1$, we have $\overline{\Delta}_1^{\mathrm{WQ}} = \frac{1}{\lambda} + \frac{1}{\mu}$; the expression coincides with that of an M/M/1/1 system with preemption reported in [72]. This is in accordance with the fact that the stream with the highest priority is not affected by any other stream.

where:

$$\overline{v}_{00} = \frac{1}{\mu} + \sum_{j=0}^{i-1} \frac{\lambda^j (i-1)!}{(i-1-j)! \sum_{k=0}^{i-1} \frac{\lambda^k (i-1)!}{\mu^k (i-1-k)!} \prod_{h=0}^{j} a_h}, \tag{3.25}$$

with $a_h \in \mathbb{R}$ is a real sequence defined as:

$$a_h = \begin{cases} \lambda + \mu, & h = i-1 \neq 0, \\ (i-h)\lambda + \mu - \frac{(i-1-h)\lambda\mu}{a_{h+1}}, & 1 \leq h \leq i-2, \\ i\lambda - \frac{(i-1)\lambda\mu}{a_1}, & h = 0, \end{cases}$$

and:

$$\overline{v}_{k0} = \frac{(i-k)\lambda}{\mu} \overline{v}_{(k-1)0} + \sum_{j=k}^{i-1} \frac{\lambda^j (i-1)!}{\mu (i-1-j)! \sum_{k=0}^{i-1} \frac{\lambda^k (i-1)!}{\mu^k (i-1-k)!}}, \quad 1 \leq k \leq i-1. \tag{3.26}$$

Proof. With the stationary distribution of the Markov chain being found, we start by applying Theorem 3.1. To properly find the vectors $\overline{\boldsymbol{v}}_q = [\overline{v}_{q0}, \overline{v}_{q1}]$, $\forall q \in \mathbb{Q}$, we proceed by applying the aforementioned theorem in the states where the service transitions with rate μ have no effect on the age process $\boldsymbol{x}(t)$ (i.e., $\boldsymbol{A_l} = \boldsymbol{I}_2$). More specifically, we start by applying Theorem 3.1 in the state $q = i-1$ and go backward in the discrete state space. By doing so, and by focusing on the first component of the vector $\overline{\boldsymbol{v}}_{i-1}$, we end up with the following:

$$\overline{v}_{(i-1)0} = \frac{\overline{\pi}_{i-1}}{\mu} + \frac{\lambda}{\mu} \overline{v}_{(i-2)0}. \tag{3.27}$$

By doing a successive backward induction till state $k = 1$, we can verify that for all states $1 \leq k \leq i-1$, we have:

$$\overline{v}_{k0} = \sum_{j=k}^{i-1} \frac{\overline{\pi}_j}{\mu} + \frac{(i-k)\lambda}{\mu} \overline{v}_{(k-1)0}. \tag{3.28}$$

The next step consists of formulating the results of Theorem 3.1 for the second component of the vector $\overline{\boldsymbol{v}}_q$. To that end, we can see that:

$$\overline{v}_{(i-1)1} = \frac{\overline{\pi}_{i-1}}{\mu + \lambda} + \frac{\lambda}{\mu + \lambda} \overline{v}_{(i-2)1}. \tag{3.29}$$

In the state $k = i-2$, and by taking into account the previous equation (3.29), we end up with:

$$\overline{v}_{(i-2)1} = \frac{\mu \overline{\pi}_{i-1}}{(\mu + \lambda)(2\lambda + \mu - \frac{\lambda\mu}{\lambda+\mu})} + \frac{\overline{\pi}_{i-2}}{2\lambda + \mu - \frac{\lambda\mu}{\lambda+\mu}} + \frac{2\lambda}{2\lambda + \mu - \frac{\lambda\mu}{\lambda+\mu}} \overline{v}_{(i-3)1}. \tag{3.30}$$

Therefore, we carefully employ a backward induction to conclude the following closed-form for all states $1 \leq k \leq i-1$:

$$\overline{v}_{k1} = \sum_{j=k}^{i-1} A_k^j + \frac{(i-k)\lambda}{a_k} \overline{v}_{(k-1)1}, \tag{3.31}$$

where $a_k \in \mathbb{R}$ is a real sequence that is defined for $1 \leq k \leq i-1$ as follows:

$$\begin{aligned} a_{i-1} &= \lambda + \mu, \\ a_k &= (i-k)\lambda + \mu - \frac{(i-1-k)\lambda\mu}{a_{k+1}}, \quad 1 \leq k \leq i-2, \end{aligned} \tag{3.32}$$

and A_k^j is equal to $\frac{\mu^{j-k}\overline{\pi}_j}{\prod_{h=0}^{j-k} a_{k+h}}$. With $\overline{v}_{k0}$ and $\overline{v}_{k1}$ for all $1 \leq k \leq i-1$ at our disposal, we continue by formulating the equations at state $k = 0$ where the transition of rate μ induces resets in the age processes:

$$\overline{v}_{00}((i-1)\lambda + \mu) = \overline{\pi}_0 + \mu\overline{v}_{01} + \mu\overline{v}_{10} \tag{3.33}$$

By using the general expression in eq. (3.28), we can conclude that:

$$\overline{v}_{00} = \frac{1}{\mu} + \overline{v}_{01}. \tag{3.34}$$

As the goal is to find $\sum_{k=0}^{i-1} \overline{v}_{k0}$, we proceed with calculating $\overline{v}_{01}$:

$$\overline{v}_{01}(i\lambda) = \overline{\pi}_0 + \mu\overline{v}_{11}. \tag{3.35}$$

By using the general expression in eq. (3.31), we can conclude that:

$$\overline{v}_{01} = \sum_{j=0}^{i-1} A_0^j, \tag{3.36}$$

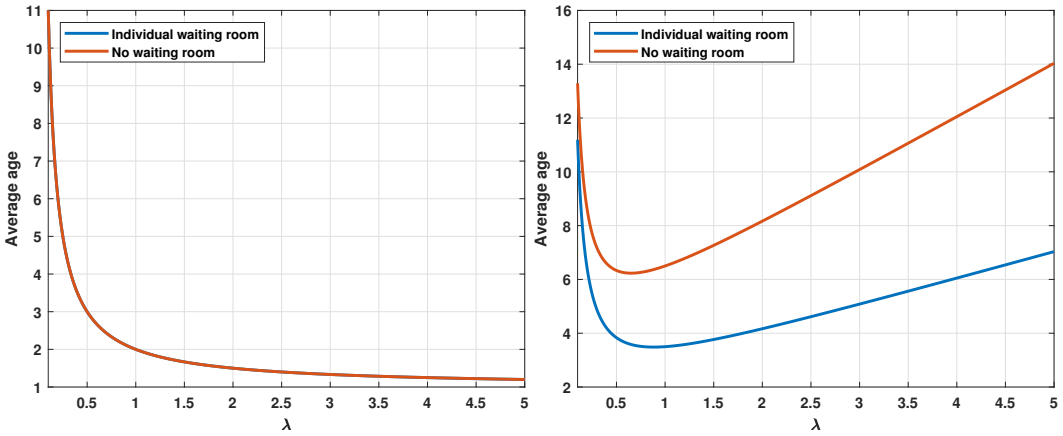
where $a_0 = i\lambda - \frac{(i-1)\lambda\mu}{a_1}$. All in all, we can conclude that:

$$\overline{v}_{00} = \frac{1}{\mu} + \sum_{j=0}^{i-1} A_0^j = \frac{1}{\mu} + \sum_{j=0}^{i-1} \frac{\lambda^j (i-1)!}{(i-j)! \sum_{k=0}^{i-1} \frac{\lambda^k (i-1)!}{\mu^k (i-1-k)!} \prod_{h=0}^{j} a_h}. \tag{3.37}$$

Knowing that $\overline{\Delta}_i^{\mathrm{WQ}} = \sum_{k=0}^{i-1} \overline{v}_{k0}$ and by taking into account the results of eq. (3.28) and (3.37), we conclude our proof. □

Numerical Implementations

In this section, we compare each stream's average in the case where a waiting room exists for each stream (referred to as WQ), and the latter case where there is no waiting room (referred to as NQ). The average age in the latter can be found in [28]. We consider in the following that $\mu = 1$ and $N = 3$. We can see in Fig. 3.4a that stream 1, in both cases, achieves the same average age since, in both cases, the highest priority stream sees the system as a preemptive M/M/1/1 system. For the second stream, as seen in Fig. 3.4b, having a waiting room clearly helps for all λ. This is because if a packet for stream 2 is being served when a packet for stream 1 arrives, it is not discarded after preemption and is, therefore, resumed right after.



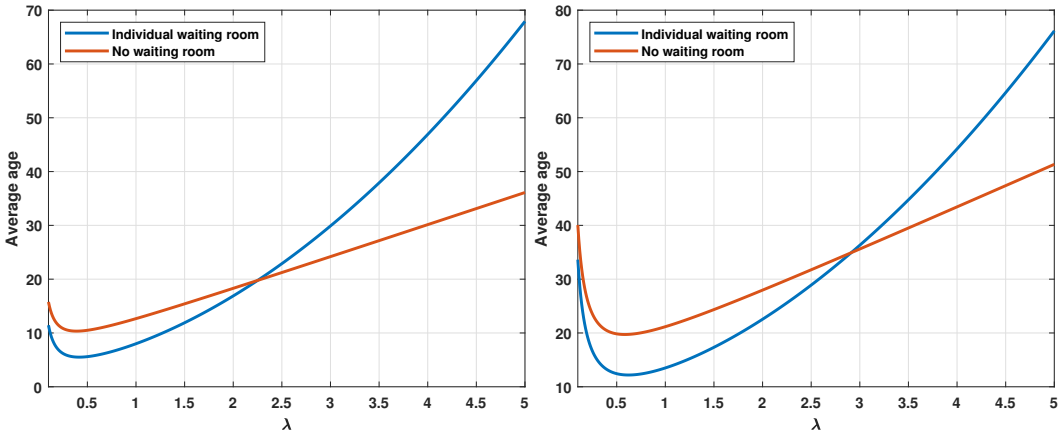
(a) Average age of stream 1.

(b) Average age of stream 2.

Figure 3.4: The average age of stream 1 and 2 in function of the arrival rate λ.

An interesting observation can be seen in Fig. 3.5a for stream 3: although a preempted packet of stream 3 is not discarded, yet this stream only exhibits an advantage compared to NQ for low λ. As λ increases, the gap gets smaller until both curves intersect, and having a waiting room worsen its performance in terms of average age. This is because although the packets of stream 3 are not being discarded after preemption, the same thing is happening for stream 2. Therefore, packets of stream 3 have to wait for the preempted packet of stream 2, by stream 1, to continue service before it can be served itself. Due to this observation, we can see in Fig. 3.5b that the total average age with a waiting room is smaller than its counterpart when $\lambda < \lambda_{PASS}$ and is higher otherwise. λ_{PASS} denotes the arrival rate corresponding to the intersection of the two curves. The value of λ_{PASS} can be found by using the results of Theorem 3.2 and that of [28].

To further highlight this trade-off, we vary the number of streams N and report the following table's results. We can notice that, in both cases, the lowest total average age is attained for the same arrival rate λ_{OPT}. However, the total average age $\overline{\Delta}^{\text{WQ}}_{\text{OPT}}$ is always smaller than its counterpart $\overline{\Delta}^{\text{NQ}}_{\text{OPT}}$ with the gap between them increasing as N grows. It is worth noting that the minimum total average age is



(a) Average age of stream 3. (b) Total average age of the system.

Figure 3.5: The average age of stream 3 and the total average age in function of the arrival rate λ.

achieved for smaller values of λ as N grows higher due to the congestion. As for the trade-off, we can see that, as N grows, λ_{PASS} becomes smaller and buffering packets instead of discarding them after preemption worsen the performance in terms of total average age when $\lambda > \lambda_{\text{PASS}}$.

N	$\overline{\Delta}^{\text{WQ}}_{\text{OPT}}$	$\lambda^{\text{WQ}}_{\text{OPT}}$	$\overline{\Delta}^{\text{NQ}}_{OPT}$	$\lambda^{\text{NQ}}_{\text{OPT}}$	λ_{PASS}
3	12.18	0.62	19.71	0.62	2.92
5	33	0.3	55	0.3	0.7
8	81.7	0.16	140	0.16	0.31

Table 3.2: Comparison between WQ and NQ scenarios.

3.4 Lexicographic Optimality for Age Minimization

The previous section's numerical results raise an essential question for multi-class settings: how can we define age-optimality? The standard total average age, or similar optimality definitions abundant in standard multi-stream systems [33], do not hold here. The reason is that the distinction in priority between the various streams typically reflects the importance of the data carried by these streams and their time-cruciality. In other words, a policy might be better to be adopted, even though it leads to a higher total average age, as long as it performs well for the highest priority streams. For example, let us consider $\lambda = 4$ in the first scenario of the previous numerical implementations. As seen in Fig. 3.5b, the policy with an individual waiting room has a larger total average age compared to the no waiting room policy. On the other hand, and as seen in Fig. 3.4b, the no waiting room policy leads to a larger age for the more critical stream 2. Therefore, one can argue

that the individual waiting room policy might lead to an overall better performance, even though it has a higher total average age. In the sequel, we will introduce the notion of Lexicographic optimality for the age minimization framework, which we will refer to as the lex-age-optimality. The lex-age-optimality elegantly solves this issue and provides a new direction of age analysis in multi-class scheduling environments.

In this section, we return to the general case where J_i can be different from 1 for any $i = 1, \ldots, I$. Moreover, we consider the following class of synchronized packet generation and arrival processes.

Definition 3.5. *Intra-class Synchronized Sampling and Arrivals: The packet generation and arrival times are said to be synchronized across streams within each class, if for all classes $i = 1, \ldots, I$, 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 $j = 1, \ldots, J_i$*

$$S_n^{i,j} = S_n^i, \quad A_n^{i,j} = A_n^i. \tag{3.38}$$

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 can occur when these streams are synchronized by the same clock, e.g., in monitoring and control applications [73,74]. An example of such a scenario is a vehicular network where safety-related data (e.g., position and velocity) are generated every T time units. In contrast, other lower priority data can have different traffic patterns (e.g., updates on the traffic are generated every T' time units) [4]. Also, when $J_i = 1$ for a particular class i, the synchronization assumption within class i reduces to arbitrary packet generation and arrival processes for the class mentioned above. It is worth mentioning that our work is not restricted to any traffic arrival distribution. It can include arbitrary arrival processes where packets may arrive out of order of their generation times. In the sequel, we let

$$\mathcal{I} = \{(S_n^i, A_n^i),\ i = 1, \ldots, I, \quad n = 1, 2, \ldots\} \tag{3.39}$$

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

With the system model of this section clarified, we can now proceed with defining the lex-age-optimality.

Definition 3.6. *Lex-age-optimality: A scheduling policy $P \in \Pi$ is said to be level 1 lex-age-optimal within Π if for all $\mathcal{I}$, $p_t \in \mathcal{P}_{sym}$ and $\pi \in \Pi$*

$$[\{p_t \circ \mathbf{\Delta}_P^1(t), t \geq 0\}|\mathcal{I}] \leq_{st} [\{p_t \circ \mathbf{\Delta}_\pi^1(t), t \geq 0\}|\mathcal{I}]. \tag{3.40}$$

We let $\Pi_{lex\text{-}opt}^1 \subseteq \Pi$ denote the set of scheduling policies that are level 1 lex-age-optimal. In addition, 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 $\mathcal{I}$, $p_t \in \mathcal{P}_{sym}$ and $\pi \in \Pi_{lex\text{-}opt}^{k-1}$

$$[\{p_t \circ \mathbf{\Delta}_P^k(t), t \geq 0\}|\mathcal{I}] \leq_{st} [\{p_t \circ \mathbf{\Delta}_\pi^k(t), t \geq 0\}|\mathcal{I}], \tag{3.41}$$

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

According to (3.12), (3.40) can be equivalently expressed as

$$\mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^1_P(t), t \geq 0\})|\mathcal{I}] = \min_{\pi \in \Pi} \mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^1_\pi(t), t \geq 0\})|\mathcal{I}], \tag{3.42}$$

for all $\mathcal{I}$, $p_t \in \mathcal{P}_{\text{sym}}$, and non-decreasing functional $\phi : \mathbb{V} \mapsto \mathbb{R}$, provided that the expectations in (3.42) exist. Similarly, an equivalent formulation of the level k lex-age-optimality (3.41) of a policy $P \in \Pi^{k-1}_{\text{lex-opt}}$ is

$$\mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^k_P(t), t \geq 0\})|\mathcal{I}] = \min_{\pi \in \Pi^{k-1}_{\text{lex-opt}}} \mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^k_\pi(t), t \geq 0\})|\mathcal{I}], \tag{3.43}$$

for all $\mathcal{I}$, $p_t \in \mathcal{P}_{\text{sym}}$, and non-decreasing functional $\phi : \mathbb{V} \mapsto \mathbb{R}$, provided that the expectations in (3.43) exist.

The goal of the lex-age-optimality is to guarantee the age-optimality of high priority classes and optimize the low priority classes' age performance accordingly. To see how this is achieved, we recall from (3.42) 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 $[\{p_t \circ \mathbf{\Delta}^1(t), t \geq 0\})|\mathcal{I}]$ among all causal policies. Next, to maintain the age-optimality of the highest priority class, our attention is restricted to scheduling policies that are level 1 lex-age-optimal. We have denoted this set by $\Pi^1_{\text{lex-opt}}$. To that end, and as seen in (3.43), 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 $[\{p_t \circ \mathbf{\Delta}^2(t), t \geq 0\})|\mathcal{I}]$ among all level 1 lex-age-optimal policies. This showcases how the lex-age-optimality captures the criticality of streams since, by definition, lex-age-optimal policies grant high priority streams the best possible performance without being influenced by low priority streams. Then, while ensuring the high priority streams' age-optimality, the performance of the low priority streams is optimized.

3.4.1 Lex-Age-Optimal Policy for Exponential Service Time

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

Definition 3.7. Informative and Non-informative Packets*: Consider a packet of stream* (i, j) *that is generated at time* $S^{i,j}_n \leq t$*. The packet is said to be informative at time* t *if* $t - S^{i,j}_n < \Delta^{i,j}(t)$*; otherwise, the packet is non-informative.*

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

Definition 3.8. Preemptive Priority (PP) policy based on Informative Packets*: Among the streams with informative packets, the class of streams with the highest priority are served first. A packet in service is preempted upon the arrival of an informative packet of a higher priority stream; the preempted packet is stored back in the queue.*

Definition 3.9. *Maximum Age First (MAF) policy: Among the streams from a priority class, the stream with the maximum age is served first, with ties broken arbitrarily.*

Definition 3.10. *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 3.11. *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.*

Note that our proposed policy does not drop non-informative packets as previously proposed in the literature (e.g., [12]). Although these packets are not necessary for reducing the age, they may still be needed at the monitor in many applications (e.g., social updates). In the case of a single priority class (i.e., $I = 1$), the proposed policy coincides with the Maximum Age First, Last-Generated, First-Served (MAF-LGFS) policy proposed in [33].

By definition, our policy ensures that the service of high priority informative packets is not interrupted nor influenced by any lower priority packets. This grants critical packets the best possible service by the facility. Note that informative packets play a crucial role in our policy. In particular, the preemptive priority discipline is a dynamic priority rule based 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 class I that have informative packets. This non-trivial aspect of our policy ensures that low priority classes are provided with the best possible transmission opportunity while not affecting the high priority streams' age. These fundamental observations are crucial and will be used to establish the lex-age-optimality of the PP-MAF-LGFS policy.

Theorem 3.3 (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 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., Theorem 1 in [33]), showing these scheduling rules are sufficient for optimality is not enough in our case. In fact, at each induction step, a characterization of the exact behavior of each policy $\pi \in \Pi^k_{\text{lex-opt}}$ for $k = 1, \ldots, I$ is required. This poses several technical difficulties, which we solve in our sample-path proof by showing the necessity of the scheduling rules for level k lex-age-optimality for $k = 1, \ldots, I$. The details can be found in Appendix A.1. □

Note that when each priority class has only one stream, the intra-class synchronization assumption is always satisfied, and Theorem 3.3 holds for *arbitrarily given* packet generation and arrival times. This special case is of particular interest.

To the best of our knowledge, this is the first lex-age-optimality results for multi-class status updates. Our optimality results are general as we established them in terms of stochastic ordering of stochastic processes for all symmetric non-decreasing penalty functions and 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 was previously explained, this is of paramount importance as priority classes typically represent different applications, each with their traffic arrivals and data timeliness requirements. For example, in a particular scenario, we can be interested in minimizing the average max-age for class 1, the time-average sum-age for class 2, and the average sum-age penalty for class 3. Theorem 3.3 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.

Numerical Implementations

We consider a vehicle in a Vehicle-To-Everything (V2X) network that sends packets to either nearby vehicles or roadside units (see [4], [75] for two surveys). In the

surveys above, a list of possible packets uses cases are presented, each of which has different priorities in the network. We consider 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 Hz. We assume in our settings that the packets' generation frequency is set to 10 Hz. This class of streams has the highest priority among all data types. In our simulations, we consider 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. In our simulations, we consider that two streams belong to this class (e.g., updates concerning the vehicle). The generation frequency of these packets is set to 0.5 Hz. The priority of this class is second to the road safety class.

3. **Convenience and Entertainment Data**: The data in this class are considered to be the least crucial as they aim to provide entertainment and convenience solely for improving the quality of travel. In our simulations, we consider that two streams belong to this class, and we suppose that the generation frequency of their packets is 5 Hz.

Based on the above, we can conclude that our considered system's arrival rate is $\lambda_{tot} = 31$ packets per second. The vehicle's service facility 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 a service rate μ.

We compare our proposed policy to the preemptive MAF-LGFS[2] policy proposed in [33]. The preemptive MAF-LGFS policy schedules the stream's packet 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 ms. By doing so, we get

$$\mathbb{E}[\phi_{\text{avg}}(\{p_{\text{exceed}-\alpha} \circ \mathbf{\Delta}^1(t), t \geq 0\})] = \frac{1}{2}\sum_{j=1}^{2}\frac{1}{T}\int_0^T \Pr(\Delta^{1,j}(t) > \alpha)dt, \qquad (3.44)$$

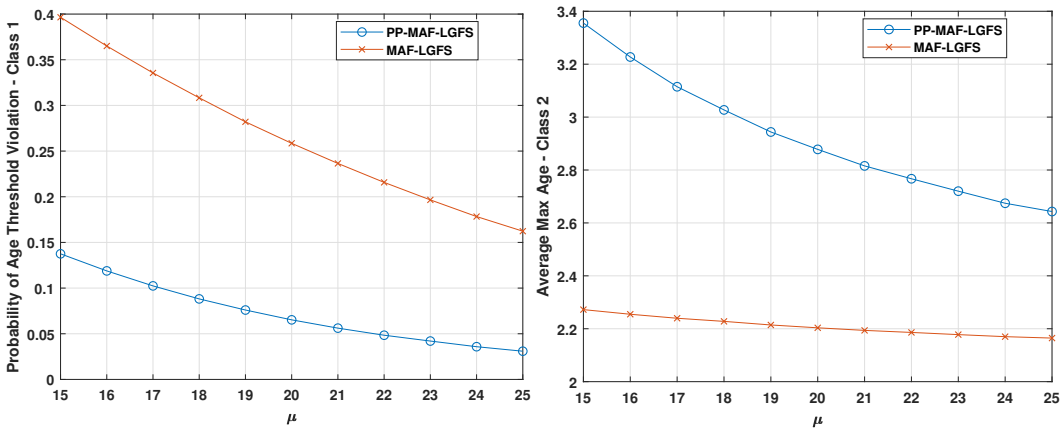
where $\Pr(\Delta^{1,j}(t) > \alpha)$ is the probability of violation of the maximum tolerated age 250 ms by stream $(1, j)$ at time t. The interest in this time-average age penalty function is that in vehicular networks, small age for the velocity and position data can be tolerated. After a certain value, the system's performance starts deteriorating due to this aging. For class 2, we choose $p_{\max}$ and ϕ_{avg} as the age penalty function and functional, respectively. In other words, we are interested in minimizing the

[2]FCFS policies are omitted from our simulations as LGFS policies will always outperform them since queuing will lead to the unnecessary staleness of the packets.

average max-age of class 2. 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 s. We report in Fig. 3.6 the simulations results that showcase the performance of each policy. We can conclude from these results the following:

- As seen in Fig. 3.6a, 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. 3.6b-3.7 that the preemptive MAF-LGFS policy outperforms our proposed policy for classes 2 and 3. In fact, in our policy, giving priority to class 1 leads to a penalty for the other classes. However, we recall that the probability of violating 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. This is because class 1's packets finish transmission much faster the higher μ is. Consequently, in our proposed policy, the server will finish serving class 1 fast enough that it can start serving the other classes before new packets for class 1 arrive at the system. This reduces the incurred penalty by the low priority classes due to the high priority streams.



(a) Probability of age violation for class 1.

(b) Average max age of class 2.

Figure 3.6: Comparison between the two policies in function of the service rate μ for classes 1 and 2.

The above results highlight our proposed lex-age-optimal policy's performance and provide a new direction on age analysis in multi-class scheduling scenarios.

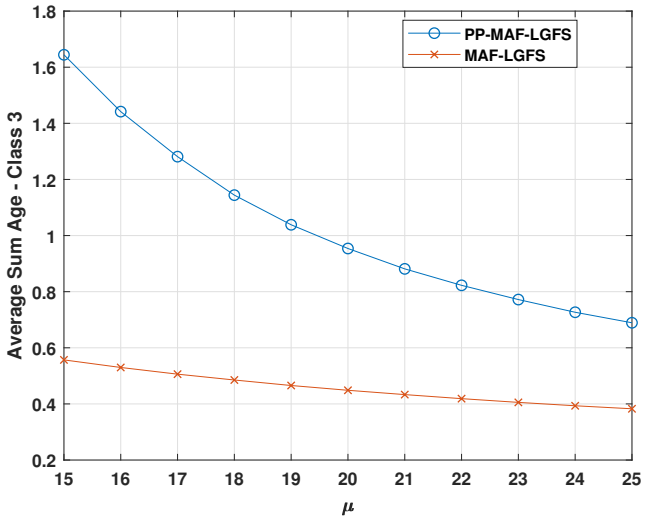


Figure 3.7: Comparison between the two policies in function of the service rate μ for class 3.

3.5 Conclusion

In this chapter, we have investigated a multi-class environment where streams are assigned different priorities. Using SHS tools, we derived a closed-form expression of the average AoI for the case of Poisson arrivals and exponential transmission times. These expressions were used to provide insights into the interaction between the different classes. Equipped with these insights, we introduced the notion of lex-age-optimality that captures both the age-optimality and the order of criticality between the streams in a general multi-class setting. To that end, we have proposed an online scheduling policy in this general multi-class, multi-stream scheduling scenario. Using a sample-path argument, we proved the proposed policy's lex-age-optimality in the single exponential server case. The optimality results hold for any symmetric non-decreasing age penalty function and all non-decreasing age penalty functionals. Numerical results were then presented to highlight the performance of our proposed policy.

4 | Average Age Minimization in a CSMA Environment

4.1 Overview

This chapter investigates a network where N links contend for the channel using the well-known CSMA scheme. As previously discussed in the thesis' introduction, CSMA is considered one of the most popular distributed MAC schemes in wireless networks. To that end, there exists a vast amount of research results on CSMA regarding its theoretical analysis and applications. For example, throughput optimal CSMA schemes have been extensively investigated in the literature [52, 76, 77]. With the AoI metric being relatively new, the question of what is the AoI optimal CSMA scheme remains open. This motivated the work in this chapter, where we theoretically study a general CSMA environment. To that end, the following are the key contributions of this chapter:

- We first leverage the notion of SHS to model a CSMA environment in which N interfering links contend for the channel. Using these tools, we derive: 1) a closed-form expression of the average age when links generate packets at will, and 2) an upper bound of the average age when packets arrive stochastically to each link. This upper bound is shown to be generally tight and equal to the average age in specific scenarios.

- Afterward, and in both scenarios, we formulate an optimization problem to minimize the total average age of the network. Interestingly, it will be shown that the minimum average age is achieved for the same back-off duration in both the sampling and stochastic arrivals scenarios. The formulated problem is then shown to have an equivalent convex form that can be solved efficiently to find each link's optimal back-off time. Based on this optimal point, theoretical insights on the interaction between links in these CSMA settings are provided. We also provide further applications of our considered model that incorporate many realistic scenarios such as hybrid networks consisting of both throughput and age-sensitive links, and networks where each link requires an average age guarantee. Following that, the proposed optimized

CSMA scheme is implemented in a realistic IEEE 802.11 setting, and its performance is highlighted. Moreover, we compare the performance of our proposed age-optimal CSMA scheme with the throughput optimal one. This comparison sheds light on essential differences between the two frameworks.

- Next, to further improve the network's performance, we propose a *novel* modification to the CSMA scheme in question. More precisely, we let each link have the freedom to transition to SLEEP mode upon successful packet transmission. To showcase the benefits of our proposed modified CSMA scheme, we first thoroughly study a simplified version of it where we find a closed-form of the average age of the network as a function of the performance parameters. This simplified version helps us capture the essence of our proposed scheme, provide essential insights on the interaction between the links, and showcase the average age gain compared to the optimized standard CSMA. Lastly, to circumvent the intractability of finding a closed-form expression of the total average age in the generic version of the proposed scheme, a Sequential Convex Approximation (SCA) approach is presented to optimize the age performance. The convergence of the SCA procedure to a stationary point is then proven. It is worth noting that our approach applies to any general average age minimization problem modeled through the same SHS tools. Simulation results are then laid out to highlight the performance gain offered by our SCA approach compared to that of the optimized standard CSMA.

The rest of the chapter is organized as follows: Section 4.2 presents the system model. Section 4.3 provides the closed-forms of the average age in both the sampling and the stochastic arrivals scenarios. Section 4.4 puts into perspective our optimization problem and our proposed solution to find the optimal operating point. Section 4.5 revolves around our newly proposed CSMA scheme with the aim of further improving the performance of the network. Numerical results that corroborate the theoretical findings are also presented within Sections 4.4 and 4.5 while Section 4.6 concludes the chapter.

4.2 System Model

We consider in this chapter N links (transmitter-receiver pairs) sharing the medium of transmission. The transmitter side of each link sends status updates to its corresponding monitor. However, due to interference, only one link can be active at each time instant. We suppose that the transmission time of the packets of link k is an exponentially distributed random variable with an average channel holding time of $\frac{1}{H_k}$. This chapter focuses on a network where links employ CSMA to access the channel. In this class of schemes, a transmitter listens to the medium before sending a packet. More specifically, a transmitter waits for a specific duration before transmitting, called the *back-off* time. While waiting, it keeps sensing

the environment to spot any conflicting transmission. If any interfering transmission is spotted, the transmitter stops its back-off timer immediately and waits for the medium to be free to resume it. In other words, the back-off timers of all links only tick when the channel is idle. After successful transmission by a particular link, the transmitter side of this link generates a new back-off time to prepare for the next packet transmission. We suppose that the chosen back-off time by any link k is exponentially distributed[1] with an average of $\frac{1}{R_k}$. It is worth mentioning that the adoption of exponential distribution for the back-off times has been widely done in the CSMA literature [77] [52]. In the following, we adopt the standard idealized CSMA assumptions [77]: 1) the problem of *hidden nodes* does not exist (i.e., all transmitters are within the same communication range), and 2) sensing is considered instantaneous. The first condition is satisfied in realistic scenarios if the range of carrier-sensing is large enough [78]. The second condition lets us capture the essence of the scheduling problem in question without being concerned about the contention resolution. We will discuss in Section 4.4 how the present work and analysis can be extended to the case where contention resolution is involved, and we will provide implementation considerations in IEEE 802.11 networks.

The next aspect of our model that we tackle is the packets' arrivals of each link in the network. We will consider in this chapter both the generate-at-will (sampling) and stochastic arrivals scenarios previously depicted in the introduction.

1. **Sampling**: In this framework, links are able to generate information updates anytime. To that end, we consider in this scenario that once a link captures the medium, it samples its process and proceed to the transmission stage. We also forgo queuing in the network since, as it was shown in [7], queues will induce unnecessary staleness to the sampled packets. It is worth noting that, in general, the time required to generate a sample is much smaller than the transmission time. Consequently, to keep our analysis tractable and as it has been previously adopted in the literature, we assume that the sampling time is negligible [7].

2. **Stochastic arrivals**: In this framework, packets are assumed to arrive randomly to each link. Subsequently, we suppose that the packet's arrival at each transmitter k is exponentially distributed with rate λ_k. Also, in this case, we forgo queuing; each transmitter keeps at most one packet in its system. Upon a new arrival of a packet to transmitter k, the packet of transmitter k that is currently available (or being served) is preempted and discarded. This setting is motivated by the fact that a preemptive M/M/1/1 scenario was shown to minimize the average age in the case of exponential transmission time [71].

Remark 4.1. *We adopt a zero-wait sampling policy in the sampling scenario. More specifically, the transmitter side of the link that captures the channel generates*

[1]The validity of the conclusions drawn from our subsequent analysis will be verified for more general back-off and service time distributions in Section 4.4.4.

a packet immediately, and the transmission phase begins. We avoid any waiting mechanisms before sampling after capturing the channel for the following reasons:

1. *The transmitters already wait a certain amount of time before transmitting as the CSMA framework requires a back-off mechanism to be adopted by each link.*

2. *As we are interested in a random access environment, any waiting before sampling by a particular link after capturing the channel will lead to unnecessary aging of the remaining links that are waiting for the channel to be free again.*

Assumption 4.1. *We assume that, in the stochastic arrivals scenario, a transmitter that captures the medium sends a "fake" update (i.e., a packet with the same timestamp as the previously transmitted packet) if its buffer is empty [64].*

The above assumption will make the mathematical model tractable. Investigating the exact case requires tracking each link's buffer status, which makes the system dimensions grow exponentially. Therefore, the performance achieved in our model constitutes an upper bound for the average age when compared to the case where only links with available packets compete for the channel. It is worth noting that the gap between the two scenarios vanishes as the traffic intensity increases. Moreover, simulation results are presented in Section 4.4.4 to showcase the tightness of this upper bound even in low traffic scenarios.

As we will investigate the AoI in both of the above frameworks, we emphasize on the wide scope of the results of this chapter as they encompasses a large variety of AoI applications. In both scenarios, we denote the instantaneous age of information at the receiver (monitor) of link k at time instant t by $\Delta_k(t)$. Each link's age will depend on the time spent to capture the medium and its transmission time. The ultimate goal therefore consists of minimizing the total average age that is defined as:

$$\overline{\Delta} = \sum_{k=1}^{N} \overline{\Delta}_k = \sum_{k=1}^{N} \lim_{T\to+\infty} \frac{1}{T} \int_0^T \Delta_k(t)dt. \tag{4.1}$$

4.3 Average Age Closed-Forms

To simplify the average age calculations, we forgo studying the age process of all links simultaneously. Instead, we examine a link of interest i and calculate its average age by considering the network from its perspective. Afterward, the total average age of the network can be easily concluded by summing over all the links. By proceeding in this direction, the application of Theorem 3.1 becomes easier as the dimensions of the following quantities will be reduced: $\boldsymbol{x}(t), \overline{\boldsymbol{v}}_q, \boldsymbol{b}_q \in \mathbb{R}^2$, $\forall q \in \mathbb{Q}$ and $\boldsymbol{A}_l \in \mathbb{R}^2 \times \mathbb{R}^2$, $\forall l \in \mathbb{L}$. In this case, we define the discrete states set $\mathbb{Q} = \{0, 1, 2, \ldots, N\}$ where $q(t) = k$ if link k has captured the channel and started transmission at time t while $q(t) = 0$ if the channel is idle at time t. The continuous-time

state process is defined as $\boldsymbol{x}(t) = [x_0(t), x_1(t)]$ where $x_0(t)$ is the age of the link of interest i at the monitor at time t and $x_1(t)$ is the age of the packet in the system of link i at time t. Our goal becomes to solve eqs. (3.6) to identify the vectors $\overline{\boldsymbol{v}}_q = [\overline{v}_{q0}, \overline{v}_{q1}]$, $\forall q \in \mathbb{Q}$. This will allow us to calculate the average age of the link of interest i through Theorem 3.1. We now proceed with investigating the network from link i's perspective in the two previously mentioned scenarios.

Sampling

To analyze this scenario, we first summarize the transitions between the discrete states and the resets they induce on the age process $\boldsymbol{x}(t)$ from the perspective of link i in the following table:

l	$q_l \to q_l'$	$\lambda^{(l)}$	$\boldsymbol{x}\boldsymbol{A}_l$	$\boldsymbol{A}_l$	$\overline{\boldsymbol{v}}_{q_l}\boldsymbol{A}_l$
1	$0 \to 1$	R_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{00}, \overline{v}_{01}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$0 \to N$	R_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{00}, \overline{v}_{01}]$
$N+1$	$1 \to 0$	H_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{10}, \overline{v}_{11}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$N+i$	$i \to 0$	H_i	$[x_1, 0]$	$\left(\begin{smallmatrix}0 & 0\\ 1 & 0\end{smallmatrix}\right)$	$[\overline{v}_{i1}, 0]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2N$	$N \to 0$	H_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{N0}, \overline{v}_{N1}]$

Table 4.1: Sampling scenario SHS description.

In the sequel, we elaborate on the transitions reported in Table 4.1:

1. The first set of transitions spanning from $l = 1$ till $l = N$ corresponds to the case where link k captures the channel. Thanks to the memoryless property of the back-off times, the rate of each transition $l = k$ is R_k. Capturing the channel does not change the age of the link of interest i at the monitor nor the age of the packet in its system. Therefore, the age process vector stays the same without any reset $\boldsymbol{x}' = \boldsymbol{x}\boldsymbol{A}_l = \boldsymbol{x}$.

2. The transitions $l = N + k$, $\forall k$ correspond to the case where link k releases the channel upon successful transmission. As can be seen, successful transmission of any packet belonging to links $k \neq i$ will not result in any reset of the age process $\boldsymbol{x}$. On the other hand, for the transition $l = N + i$, the age at the monitor of link i resets to the age of the packet that was delivered x_1. As there are no more packets in the system for link i, the age x_1' is set to 0.

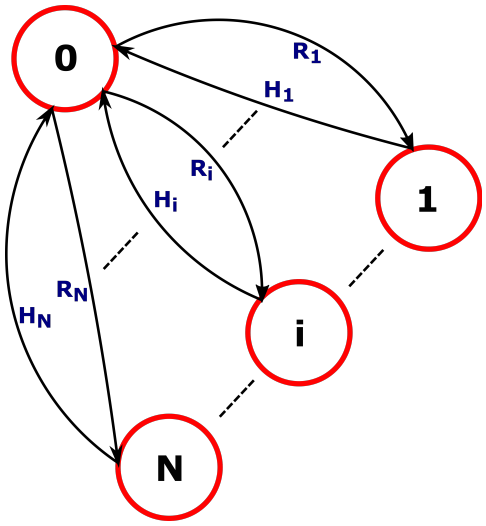


Figure 4.1: Illustration of the stochastic hybrid systems Markov chain for the sampling scenario.

As for the differential equations governing the evolution of the age process, we know that the age at the monitor always increases at a unit rate (i.e., $\dot{x}_0(t) = 1, \forall q \in \mathbb{Q}$). On the other hand, we recall that link i samples its process only when it captures the channel. In other words, there are no packets in the system for link i except in state $q(t) = i$. Consequently, $x_1(t)$ increases at a unit rate solely when $q(t) = i$:

$$\boldsymbol{b}_q = [1 \;\; 0], \quad \forall q \neq i, \tag{4.2}$$

$$\boldsymbol{b}_q = [1 \;\; 1], \quad q = i. \tag{4.3}$$

To proceed with applying Theorem 3.1, we first have to find the stationary distribution of the Markov Chain that models the transitions reported in Table 4.1 and shown in Fig. 4.1. For this purpose, we provide the following proposition.

Proposition 4.1. *The continuous-time Markov chain is irreducible, time-reversible, and admits $\overline{\pi}_k$ as stationary distribution for any feasible state $0 \leq k \leq N$ where*

$$\begin{cases} \overline{\pi}_0 = \frac{1}{C(\boldsymbol{R})}, \\ \overline{\pi}_k = \frac{R_k/H_k}{C(\boldsymbol{R})}, \qquad k = 1, \ldots, N, \end{cases} \tag{4.4}$$

and $C(\boldsymbol{R})$ is a normalization factor that is equal to

$$C(\boldsymbol{R}) = 1 + \sum_{k=1}^{N} \frac{R_k}{H_k}. \tag{4.5}$$

Proof. It is sufficient to show that the above distribution verifies the detailed balance equations. Consider the two states 0 and k. We have from (4.4) that $\frac{\overline{\pi}_k}{\overline{\pi}_0} = \frac{R_k}{H_k}$ which is exactly the detailed balance equation between states 0 and k. This holds for any of the states k and, therefore, the proof is complete. □

Equipped with the above results, we provide the following theorem.

Theorem 4.1. *In the sampling scenario, the total average age is*

$$\overline{\Delta}(\boldsymbol{R}) = N\frac{\sum_{k=1}^{N}\frac{R_k}{H_k^2}}{C(\boldsymbol{R})} + C(\boldsymbol{R})\left(\sum_{k=1}^{N}\frac{1}{R_k}\right). \tag{4.6}$$

Proof. The proof is based on solving eqs. (3.6) and applying Theorem 3.1, along with taking into account both the transitions of Table 4.1 and the differential equations vector of eqs. (4.2) and (4.3). More specifically, we start by solving eqs. (3.6) for all states $k \neq 0$:

$$\overline{v}_{k0} = \frac{\overline{\pi}_k}{H_k} + \frac{R_k\overline{v}_{00}}{H_k}, \quad \forall k, \tag{4.7}$$

$$\overline{v}_{k1} = \frac{R_k\overline{v}_{01}}{H_k}, \quad \forall k \neq i, \tag{4.8}$$

$$\overline{v}_{i1} = \frac{\overline{\pi}_i}{H_i} + \frac{R_i\overline{v}_{01}}{H_i}. \tag{4.9}$$

By solving eqs. (3.6) at state 0, we have that:

$$\overline{v}_{00}\left(\sum_{k=1}^{N}R_k\right) = \overline{\pi}_0 + \sum_{\substack{k=1\\k\neq i}}^{N}H_k\overline{v}_{k0} + H_i\overline{v}_{i1}, \tag{4.10}$$

$$\overline{v}_{01}\left(\sum_{k=1}^{N}R_k\right) = \sum_{\substack{k=1\\k\neq i}}^{N}H_k\overline{v}_{k1}. \tag{4.11}$$

By combing the results of eqs. (4.8) and (4.11), we can show that $\overline{v}_{01} = 0$ and consequently $\overline{v}_{k1} = 0$, $\forall k \neq i$ and $\overline{v}_{i1} = \frac{\overline{\pi}_i}{H_i}$. An intuitive reason why $\overline{v}_{01} = \overline{v}_{k1} = 0$, $\forall k \neq i$ is the fact that link i samples its process only when it captures the channel. In other words, whenever $q \neq i$, there is no packet in the system for link i and the age x_1 is therefore null[2]. By taking these results into account, and by substituting eqs. (4.7) in eq. (4.10), we get that $\overline{v}_{00} = \frac{1}{R_i}$. By replacing $\overline{v}_{00}$ in (4.7) and by noting that $\overline{\Delta}_i(\boldsymbol{R}) = \overline{v}_{00} + \sum_{k=1}^{N}\overline{v}_{k0}$, we end up with[3]:

$$\overline{\Delta}_i(\boldsymbol{R}) = \frac{C(\boldsymbol{R})}{R_i} + \frac{\sum_{k=1}^{N}\frac{R_k}{H_k^2}}{C(\boldsymbol{R})}. \tag{4.12}$$

As these results are general for any link i, we can conclude the total average age by merely summing over all links. □

[2]This argument will be used in subsequent proofs throughout the chapter.

[3]The non-negativity of $\overline{v}$ is straightforward and the SHS is therefore stable.

Stochastic Arrivals

By employing the same methodology of the previous scenario, we present the following results.

Theorem 4.2. *In the stochastic arrivals scenario, and under Assumption 4.1, the total average age is*

$$\overline{\Delta}(\boldsymbol{R}) = \sum_{k=1}^{N} \frac{1}{\lambda_k} - \sum_{k=1}^{N} \frac{1}{H_k} + N \frac{\sum_{k=1}^{N} \frac{R_k}{H_k^2}}{C(\boldsymbol{R})} + C(\boldsymbol{R}) \left(\sum_{k=1}^{N} \frac{1}{R_k} \right) \tag{4.13}$$

Proof. The proof can be found in Appendix B.1. □

Remark 4.2. *An important observation we notice in eq. (4.13) is the fact that the total average age depends on the arrival rate of each link solely through the term* $\sum_{k=1}^{N} \frac{1}{\lambda_k}$. *This means that the optimization of the average back-off times in this scenario is* independent *of the arrival rate as the optimization problem can be solved without the additive term* $\sum_{k=1}^{N} \left(\frac{1}{\lambda_k} - \frac{1}{H_k} \right)$. *This makes the practical implementation of the proposed CSMA scheme interesting as there is no need to have any prior knowledge on the arrival rate of each link to calibrate their average back-off time.*

Remark 4.3. *Another interesting conclusion we can draw from the two expressions in both Theorems 4.1,4.2, and the previous remark is that the same back-off rate* $\boldsymbol{R}^*$ *minimizes both average age expressions. This observation is extremely valuable in our considered CSMA model because whatever the nature of the packet arrivals is, the same optimization problem has to be solved. For example, if we consider a network where* M *links generate packets at will while* $L = N - M$ *links receive packets stochastically, it is sufficient to optimize the following function to minimize the total average age:*

$$\overline{F}(\boldsymbol{R}) = N \frac{\sum_{k=1}^{N} \frac{R_k}{H_k^2}}{C(\boldsymbol{R})} + C(\boldsymbol{R}) \left(\sum_{k=1}^{N} \frac{1}{R_k} \right). \tag{4.14}$$

4.4 Average Age Optimization

Before we formulate the average age minimization problem, we first elaborate on our assumption of zero sensing delay. The sensing time is typically negligible compared to the transmission time, which makes this assumption viable. However, this assumption, along with the continuous nature of the back-off time, makes collisions

impossible. In practice, the sensing delay cannot be neglected and, therefore, the back-off time is chosen as a multiple of mini-slots. The duration of the mini-slot T_{slot} is dictated by physical limitations such as propagation delay (the time necessary for the receiver to detect the radio signals). Consequently, collisions can happen due to the discrete nature of the back-off times. Accordingly, for our proposed analysis to hold, we need to investigate this practical issue. In the following, we elaborate on how our collision-free theoretical analysis from the previous sections can be extended to the case where collisions may occur by simply adding a constraint to our average age minimization problem.

4.4.1 Contention Resolution

As previously mentioned, the back-off time of each link is chosen in practice as a multiple of mini-slots, each of duration T_{slot}. More specifically, each link k picks a random back-off window uniformly distributed from the range $[0, W_k - 1]$ where W_k is referred to as the Contention Window (CW) of link k. In this case, the average back-off time becomes $T_{slot}\frac{W_k-1}{2}$. Knowing that our theoretical analysis assumes an average back-off time of $\frac{1}{R_k}$, the following relationship always holds:

$$T_{slot}\frac{W_k - 1}{2} = \frac{1}{R_k}. \tag{4.15}$$

To ensure that our collision-free theoretical analysis holds in this case, we aim to upper bound the collisions probability p_c by a maximum allowed collisions probability p_{max}. The quantity p_{max} is chosen to be sufficiently small to ensure that collisions are rare in the network. To achieve this goal, we propose to lower bound the contention window of each link by a minimum contention window W_0 (i.e., $W_k \geq W_0, \ \forall k$). Simply put, each link will be forced to use a contention window that is larger or equal to W_0. Note that the probability of collisions increases the lower the contention windows of links are. Therefore, we have that the probability of collisions p_c is upper bounded by p_{cu}, where p_{cu} refers to the collisions probability when all links adopt the same minimum allowed contention window W_0. In this case, and by leveraging the results proposed by Bianchi [79], we have:

$$p_{cu} = 1 - (1 - \tau)^{N-1}, \tag{4.16}$$

where τ is the transmission probability of each link and is equal to $\frac{2}{W_0+1}$ [79]. As $p_c \leq p_{cu}$, then to guarantee that the probability of collisions is less than p_{max}, it is sufficient to upperbound p_{cu} by p_{max}. Specifically:

$$1 - (1 - \tau)^{N-1} \leq p_{max}. \tag{4.17}$$

After algebraic manipulations, we get the following condition:

$$\tau \leq 1 - \exp\left(\frac{\log(1 - p_{max})}{N - 1}\right). \tag{4.18}$$

As W_0 is the minimum contention window, we can conclude from (4.15) that the highest back-off rate R_k that can be used by any link k is:

$$R_{UB} = \frac{2}{(W_0 - 1)T_{slot}}. \tag{4.19}$$

Consequently, by using the above results, and as τ is equal to $\frac{2}{W_0+1}$ in this case, we can conclude that the condition in (4.18) is equivalent to the following:

$$R_{UB} \leq \frac{\frac{1}{T_{slot}}}{\frac{1}{1-\exp\left(\frac{\log(1-p_{max})}{N-1}\right)} - 1}. \tag{4.20}$$

Accordingly, in order to have a probability of collisions less than p_{max}, it is sufficient to upper bound the back-off rate of each link by a constant $R_{UB} = \frac{\frac{1}{T_{slot}}}{\frac{1}{1-\exp\left(\frac{\log(1-p_{max})}{N-1}\right)}-1}$. By letting p_{max} be small enough that collisions can be ignored, our collision-free theoretical analysis will still hold in this case. Therefore, it suffices to minimize the average age while considering that a back-off rate upper bound cannot be surpassed. After optimizing the back-off rate of the links subject to this constraint, each link's optimal contention window can be deduced using (4.15). Implementations of this approach in a realistic IEEE 802.11 environment will be presented in Section 4.4.4, and it will be shown that the performance degradation in comparison to that of the idealized collision-free CSMA settings is minor. The question that remains, and that the next section will answer is how to optimize the back-off rate of each link?

4.4.2 Average Age Minimization

The ultimate goal of the work is to optimize the average back-off time of each link in a way to minimize the total average age of the network. As pointed out in Remark 4.3, the total average age of the network is minimized by the same optimal back-off rate $\boldsymbol{R}^*$ in both the sampling and the stochastic arrivals scenarios. Therefore, we focus in the following on the sampling scenario case while bearing in mind that the difference between the two scenarios lies solely in the additive term $\sum_{k=1}^{N} \frac{1}{\lambda_k} - \sum_{k=1}^{N} \frac{1}{H_k}$. Consequently, our optimization problem is as follows:

$$\begin{aligned} \underset{\boldsymbol{R}}{\text{minimize}} \quad & N\frac{\sum_{k=1}^{N} \frac{R_k}{H_k^2}}{C(\boldsymbol{R})} + C(\boldsymbol{R})\left(\sum_{k=1}^{N} \frac{1}{R_k}\right), \\ \text{subject to} \quad & 0 \leq R_k \leq R_{UB}, \qquad k = 1, \ldots, N. \end{aligned} \tag{4.21}$$

At a first glimpse, the problem in (4.21) appears to be a special case of the well-known sum of ratio problems which are generally hard to solve [80]. However, by analyzing its structure, this optimization problem can be converted into an equivalent convex problem via variable substitutions as it will be shown in the following.

To put this into perspective, let us introduce the new variable $\epsilon = \frac{1}{C(\boldsymbol{R})}$ and the variables f_k such that $f_k = \epsilon R_k, \ \forall k$. As $0 \leq R_k \leq R_{UB}, \ \forall k$, we can conclude that these constraints translate into the following: $0 \leq f_k \leq \epsilon R_{UB}, \ \forall k$. Next, we can observe that $\epsilon = \frac{1}{C(\boldsymbol{R})} = \overline{\pi}_0$ and is therefore upper bounded by 1. Afterward, as the highest possible value of R_k is R_{UB} and as $\frac{\partial \epsilon}{\partial R_k} = \frac{-1}{H_k C(R)} < 0$, we can conclude that $\epsilon \geq \frac{1}{1+\sum_{k=1}^{N} \frac{R_{UB}}{H_k}}$. Lastly, by substituting R_k with $\frac{f_k}{\epsilon}$ in the expression of ϵ, the relationship between ϵ and f_k can be captured through the following equality: $\sum_{k=1}^{N} \frac{f_k}{H_k} = 1 - \epsilon$. By combining all these observations, our optimization problem can be reformulated as follows:

$$\begin{aligned} \underset{\boldsymbol{f},\epsilon}{\text{minimize}} \quad & N \sum_{k=1}^{N} \frac{f_k}{H_k^2} + \sum_{k=1}^{N} \frac{1}{f_k}, \\ \text{subject to} \quad & 0 \leq f_k \leq \epsilon R_{UB}, \qquad k = 1, \ldots, N, \\ & \frac{1}{1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}} \leq \epsilon \leq 1, \\ & \sum_{k=1}^{N} \frac{f_k}{H_k} = 1 - \epsilon. \end{aligned} \tag{4.22}$$

We analyze the above problem in the next theorem, where we establish its convexity. This enables us to find its optimal solution, which puts into perspective the importance of our proposed transformation.

Theorem 4.3. *The optimal back-off rate of each link k is*

$$R_k^* = \begin{cases} R_{UB}, & \text{if } \mu_k^* > 0, \\ \sqrt{\frac{H_k}{\epsilon^{*2}(\frac{N}{H_k} + \rho^*)}}, & \text{if } \mu_k^* = 0, \end{cases} \tag{4.23}$$

where μ_k^*, ρ^* and ϵ^* satisfy

$$\mu_k^* = -\frac{N}{H_k^2} + \frac{1}{\epsilon^{*2} R_k^{*2}} - \frac{\rho^*}{H_k}, \qquad k = 1, \ldots, N, \tag{4.24}$$

$$\begin{aligned} \rho^* &= R_{UB} \sum_{k=1}^{N} \mu_k^*, \\ \epsilon^* &= \frac{1}{C(\boldsymbol{R}^*)}. \end{aligned} \tag{4.25}$$

Proof. The proof can be found in Appendix B.2. □

Lemma 4.1. *If* $H_k = H, \ \forall k$, *then* $R_k^* = R_{UB}, \ \forall k$.

Proof. The proof can be found in Appendix B.3. □

Remark 4.4. *The results of Lemma 4.1 put into perspective that when links have the same average channel holding time* $\frac{1}{H}$*, no priority is given to any of the links in accessing the channel. The intuition is that when links have the same channel holding time, transmission by any of the links will lead to the same average increase in the age of the other idle links. Therefore, there are no grounds for giving priority to any of the links in accessing the channel. Consequently, to minimize the total average age, all we need to do is send as many packets as possible while making sure collisions are rare (i.e.,* $R_k = R_{UB}, \forall k$*). On the other hand, as it will be highlighted in 4.4.4, this will not hold in the asymmetric average channel holding time scenario.*

4.4.3 Further Applications

Average Age Minimization With Throughput Requirements

In several realistic scenarios, the network can incorporate a variety of applications:

1. AoI-sensitive applications that require the information at the receiver to be as fresh as possible.

2. Throughput sensitive applications that necessitate a minimum throughput guarantee to ensure the seamless flow of information (e.g., queuing stability).

We investigate in the following this type of hybrid networks. More specifically, we suppose that a subset $\mathcal{L}$ of links, such that $|\mathcal{L}| = L$ with $1 \leq L < N$, demands a fixed throughput guarantee. It is worth noting that these links can generate packets at will or have stochastic arrivals. Moreover, they can adopt any queuing discipline and are not bound to the preemptive LCFS discipline employed by age-sensitive links. The rest of the links $\mathcal{M}$ with $|\mathcal{M}| = M = N - L$ require their age to be the smallest possible. Without loss of generality, we suppose that the throughput required by each link belonging to $\mathcal{L}$ is normalized with respect to the link capacity (i.e., the throughput required by link l is $\overline{s}_l < 1$) [77] [52]. In this case, $\overline{s}_l$ can be viewed as the fraction of time link l has to be transmitting. By employing our CSMA scheme, we know that the state l of the Markov chains reported in Figs. 4.1 and B.1, for the sampling and stochastic arrivals cases, respectively, corresponds to the case where link l is transmitting. Consequently, we can define the normalized throughput of each link l, denoted by s_l, as the amount of time the Markov chain is in state l. More specifically:

$$s_l = \overline{\pi}_l = \frac{\frac{R_l}{H_l}}{C(\boldsymbol{R})}. \tag{4.26}$$

Naturally, we assume that the total required throughput by all links $\mathcal{L}$ is feasible, i.e., there exist a vector $\boldsymbol{R}$ such that $0 \leq R_k \leq R_{UB}, \forall k$ and $s_l \geq \overline{s}_l, \ \forall l \in \mathcal{L}$.

Consequently, by using the results of (4.26), and by following the same change of variables as the previous section, the throughput requirement condition $s_l \geq \overline{s}_l, \forall l \in \mathcal{L}$ can be easily incorporated in our analysis by adding L simple convex constraints to the problem in (4.22) as follows:

$$\begin{aligned} \underset{\boldsymbol{f},\epsilon}{\text{minimize}} \quad & M \sum_{m=1}^{M} \frac{f_m}{H_m^2} + \sum_{m=1}^{M} \frac{1}{f_m}, \\ \text{subject to} \quad & 0 \leq f_k \leq \epsilon R_{UB}, \qquad k = 1, \ldots, N, \\ & \frac{1}{1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}} \leq \epsilon \leq 1, \\ & \sum_{k=1}^{N} \frac{f_k}{H_k} = 1 - \epsilon, \\ & f_l - H_l \overline{s}_l \geq 0, \quad \forall l \in \mathcal{L}. \end{aligned} \tag{4.27}$$

One can easily see that the obtained problem is convex in $\boldsymbol{f} = [f_1, \ldots, f_N]$ and ϵ. Therefore, the optimal point $(\boldsymbol{f}^*, \epsilon^*)$ can be easily found by invoking the same convex optimization tools that were used in the previous subsection. The optimal back-off rate of each link k can then be deduced by noting that $R_k^* = \frac{f_k^*}{\epsilon^*}, \forall k$.

Average Age Minimization With Average Age Guarantee

Another additional application is a scenario where we give each link i, belonging to a subset of links $\mathcal{I}$, a guaranteed minimum average age $\overline{a}_i$. An example is when priority is given to certain links that carry critical data and require an average age guarantee[4]. To incorporate this guarantee, and by following the same change of variables as the previous section, it is sufficient to add the simple convex constraint to the problem in (4.22) for each average age guarantee condition:

1. Sampling:

$$\overline{\Delta}_i(\boldsymbol{R}) = \frac{1}{f_i} + \sum_{k=1}^{N} \frac{f_k}{H_k^2} \leq \overline{a}_i, \quad \forall i \in \mathcal{I}. \tag{4.28}$$

2. Stochastic arrivals:

$$\overline{\Delta}_i(\boldsymbol{R}) = \frac{1}{\lambda_i} - \frac{1}{H_i} + \frac{1}{f_i} + \sum_{k=1}^{N} \frac{f_k}{H_k^2} \leq \overline{a}_i, \quad \forall i \in \mathcal{I}. \tag{4.29}$$

As these additional constraints are convex, the optimization problem remains convex. The optimal point $(\boldsymbol{f}^*, \epsilon^*)$ can be, therefore, easily found by invoking convex optimization tools.

[4]It is straightforward to assume that the age guarantees for all links are feasible, i.e., there exists a vector $\boldsymbol{R}$ such that $0 \leq R_k \leq R_{UB}, \forall k$ and $\overline{\Delta}_i(\boldsymbol{R}) \leq \overline{a}_i, \forall i \in \mathcal{I}$.

Remark 4.5. *It is worth noting that the two mentioned applications can be easily combined together through their respective constraints.*

4.4.4 Numerical Results

In the following, we provide numerical results that shed light on numerous aspects of our proposed age-optimal CSMA scheme. As pointed out in our theoretical analysis, the minimum average age is achieved for the same back-off duration in both the sampling and the stochastic arrivals scenarios. We, therefore, focus in the following, unless stated otherwise, on the sampling case.

Optimal Operating Point

In the first set of simulations, we consider 2 links with link 1 and 2 having an average channel holding time of 1ms and 0.2ms respectively. We set the contention window lower bound to $W_0 = 16$. The slot time is set to $T_s = 9\mu s$ (as adopted in IEEE 802.11n [47]), which leads to R_{UB} being equal to 14.8 (we recall the results of (4.19)). We report the total average age of the system in function of R_1 and R_2 in Fig. 4.2. The first thing we can see is that if both R_1 and R_2 approach zero, the total average age is high as links barely access the channel. Also, if only one of them approaches zero, the total average age grows rapidly due to the starvation of that link. The optimal total average age of the network (marked in red) was achieved for $(\overline{\Delta}^* = 3.64, R_1^* = 5.16, R_2^* = 14.8)$. This allows us to conclude the following.

Conclusion 4.1. *In a CSMA environment where links use exponential back-off times and have exponential service times, we should prioritize the transmission of packets from fast service rate links. We can achieve this by increasing their aggressiveness on the channel (short average back-off time). The intuition behind this is that transmission by a fast link will have a smaller impact on the age of the other links compared to transmission by a slow link.*

Validity of Conclusion 4.1 for More General Distributions

Intuitively, the above conclusion should hold for a larger class of back-off and service time distributions. To affirm our intuition, we consider 2 users contending for the channel. Instead of the exponential back-off times, we suppose that the back-off time is chosen as a multiple of slot times T_s uniformly distributed in $[0, CW - 1]$. CW is referred to as the contention window and the time slot is set to $T_s = 9\mu s$. This is in accordance with the IEEE 802.11 standard [47]. We also consider a variety of transmission time distributions:

- ***Constant time***: The transmission time of users 1 and 2 are fixed to $T_{tr_1} = 1$ms and $T_{tr_2} = 5$ms respectively.

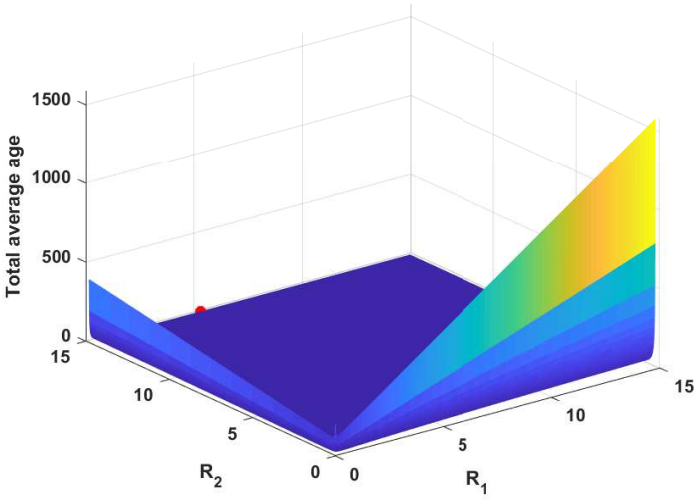


Figure 4.2: The average age in function of R_1 and R_2.

- ***Gamma distribution***: We consider that the shape parameter is equal to 2 for both users. The scale parameter is set to $\frac{1}{2}$ and $\frac{5}{2}$ for users 1 and 2 respectively. Accordingly, the average transmission time of users 1 and 2 is equal to 1ms and 5ms respectively.

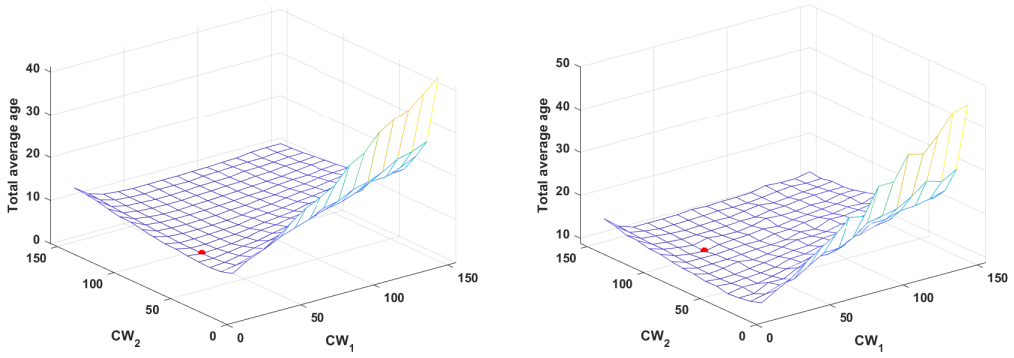
We iterate over a large space of contention window values for both users. The results of this realistic CSMA simulation are reported in Fig. 4.3a and 4.3b, where we highlight the age-optimal point in red. The optimal point is: $(\Delta^* = 7.3, CW_1^* = 26, CW_2^* = 56)$ and $(\Delta^* = 8.7, CW_1^* = 46, CW_2^* = 106)$ for the constant and gamma distributed service time respectively. In both cases, the optimal point reveals that links with fast service rates are assigned low contention windows (i.e., short average back-off time). This affirms our intuition that the insights provided by our exact analysis in the exponential case can still hold for more general back-off and service time distributions.

IEEE 802.11 Implementations

In this scenario, we put into perspective our proposed contention resolution approach presented in Section 4.4. To do so, we consider an access point communicating with N nodes and compare the performance of the two following cases:

- The unconstrained idealized CSMA settings case where collisions are assumed to be impossible. This refers to the optimal age of the convex problem in (4.22) with the imposed constraints being relaxed (i.e., $R_{UB} \longrightarrow +\infty$).

- The practical implementation of our proposed algorithm in an IEEE 802.11 environment with a slot time of $T_s = 9\mu s$. This refers to the case where: 1) R_{UB} is set based on the contention resolution approach of Section 4.4.1 and,



(a) The average age in function of CW_1 and CW_2 for constant service time.

(b) The average age in function of CW_1 and CW_2 for gamma distributed service time.

Figure 4.3: Simulation results for constant and gamma distributed service time.

2) the contention window of each link is deduced by solving (4.22) and using the relationship in (4.15). These values are disseminated to each link by the access point.

We suppose that each link in the network has an average channel holding time of 1 ms. We vary the density of the nodes and use CVX [81] to solve the corresponding optimization problems. We can see in Fig. 4.4 that the performance of our IEEE 802.11 implementations virtually coincides with the unconstrained idealized CSMA settings for small values of N. As the nodes' density increases, a gap starts forming between the realistic IEEE 802.11 implementations and the idealized CSMA settings. This is because as the density of nodes increases, collisions will have a more severe impact on the performance. However, one can see that even for a density of 8 nodes in a single collision domain, the performance degradation compared to the perfect idealized CSMA settings is only around 10%.

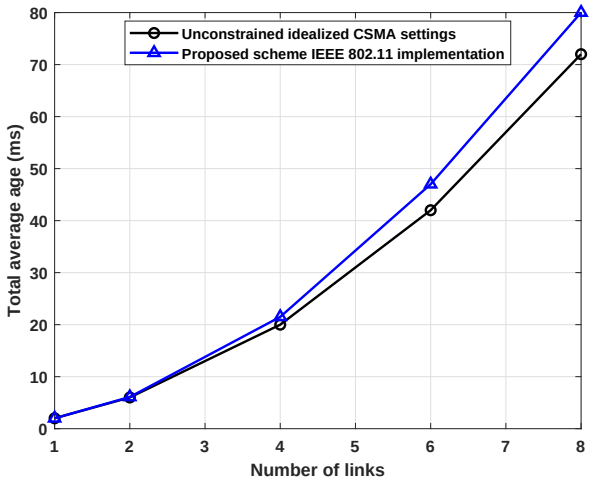


Figure 4.4: The optimal average age in function of the nodes' density.

Assumption 4.1 Verification

In the same IEEE 802.11 environment, we compare our approach for the stochastic arrival case based on the fake updates assumption to the case where links only access the channel when they have packets in their buffers. Specifically, we let “Packets CSMA" denote the optimal CSMA scheme where only links with packets compete for the channel. As we have pointed out before, our framework constitutes an upper bound for the age performance compared to the “Packets CSMA" case. However, the gap between the two performances vanishes as traffic intensity increases. For this reason, we compare the two cases for low traffic to showcase the tightness of the upper bound provided by our proposed optimization framework, even for low traffic. We suppose that $H_1 = H_2 = 1$, and we vary the arrival rates $\lambda_1 = \lambda_2 = \lambda$. The back-off rates (R_1, R_2) are optimized to minimize the total average age. One can see in Fig. 4.5 that the two plots are close, with the gap between them being 1.9 at $\lambda = 0.05$ and 0.9 at $\lambda = 1$. Beyond this point, the gap between the two plots tends to zero. This highlights that the mathematical tractability offered by our “fake" updates framework comes at a small penalty in terms of performance in low traffic scenarios while bearing in mind that this penalty disappears in high traffic scenarios.

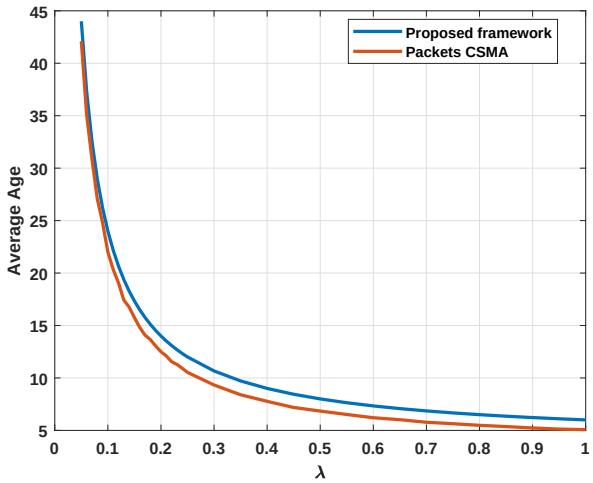


Figure 4.5: Comparison between our proposed analytical framework and Packets CSMA.

Comparison with Throughput Optimal CSMA

As it was previously reported in eq. (4.26), the total normalized throughput of the network can be expressed as

$$\overline{S} = \sum_{k=1}^{N} s_k = \frac{\sum_{k=1}^{N} \frac{R_k}{H_k}}{C(\boldsymbol{R})}, \tag{4.30}$$

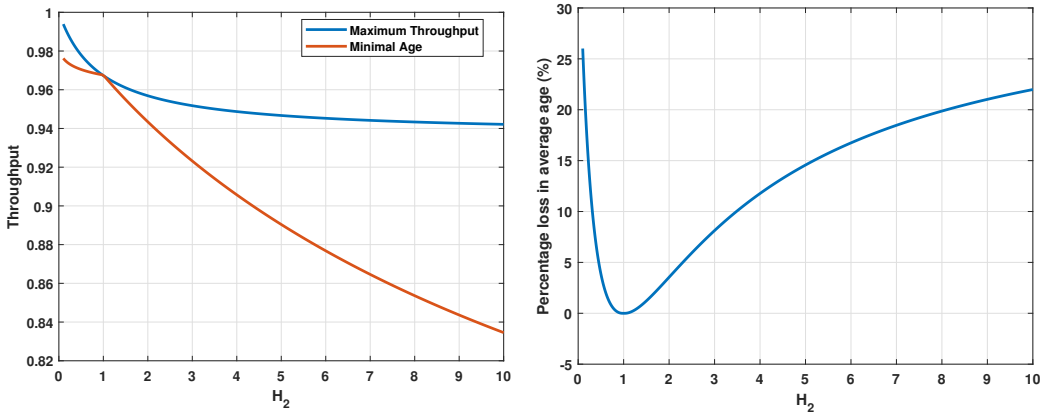
where $C(\boldsymbol{R})$ is equal to $1 + \sum_{k=1}^{N} \frac{R_k}{H_k}$. Therefore, the optimization problem of maximizing the total throughput in the network can be formulated as follows:

$$\begin{aligned} \underset{\boldsymbol{R}}{\text{maximize}} \quad & \frac{\sum_{k=1}^{N} \frac{R_k}{H_k}}{C(\boldsymbol{R})}, \\ \text{subject to} \quad & 0 \leq R_k \leq R_{UB}, \qquad k = 1, \ldots, N. \end{aligned} \tag{4.31}$$

To solve the above problem, we first rewrite $\overline{S}$ as $1 - \frac{1}{C(\boldsymbol{R})}$. Consequently, we can see that $\overline{S}$ is increasing with $\boldsymbol{R}$. Therefore, the optimal back-off rate that maximizes the total throughput of the network is $R_k = R_{UB}, \ \forall k$. The above results show that a maximum throughput policy gives the maximum allowed back-off rate to all links since the goal is to send as many packets as possible regardless of the source. On the other hand, the minimum age policy takes into account the transmission time of links, as shown in the first part of the numerical results. To compare the two policies, we consider a 2 links scenario with $H_1 = 1$, and we iterate over a range of H_2 with R_{UB} being fixed to 14.8. First of all, we compare the two policies in terms of throughput. As seen in Fig. 4.6a, the maximum throughput policy always achieves a higher throughput when compared to the minimum age policy. In fact, the minimum age policy can be seen to be far from throughput optimal. The performance of the two policies coincides when $H_1 = H_2 = 1$ since, for this case, the optimal back-off rate for the minimum age policy is $R_k = R_{UB}, \ \forall k$ (this has been theoretically proven in Lemma 4.1).
Next, we compare the two policies in terms of average age. To do so, we provide the percentage of loss in average age when the maximum throughput policy is adopted in Fig. 4.6b. The percentage of loss in average age loss is defined as the difference in average age between the two policies divided by the age-optimal policy's average age. For the same reason stated above, we can see that the two schemes' performance coincides when $H_1 = H_2 = 1$. As for the rest of the values of H_2, we can observe that the throughput optimal CSMA scheme is far from age-optimal, with the loss factor reaching as high as 25%.

4.5 Improved CSMA Scheme

To further reduce the average age of the network, we observe the following: when link i captures the channel, it samples its process and proceeds to the packet transmission stage. Upon successful transmission, the age process of the link mentioned above drops. As the monitor now has fresh knowledge of the process of link i, a new transmission right after does not necessarily provide substantial gain for the total average age of the network. More specifically, letting other links send their more needed packets sounds more beneficial to the network's overall performance. In our CSMA model of the previous sections, each link always competes for the channel whether or not it has just finished transmitting a packet. Consequently, the CSMA model needs to be modified to incorporate a new mechanism that enables



(a) Comparison in throughput between the two policies.

(b) Average age loss by the maximum throughput policy.

Figure 4.6: Comparison between the throughput and age optimal CSMA schemes.

the channel to be freed by links who have just finished transmitting their packets. To proceed in that direction, we let each link transition to SLEEP mode straight after successful transmission. In SLEEP mode, the link does not contend for the channel. Therefore, the burden on the channel is reduced to let other links have more freedom with the available resources. In the sequel, we focus on the sampling scenario while bearing in mind that the subsequent work can be easily extended to the case of stochastic arrivals. This is possible by, as it has been done in previous sections, including the arrival rate transitions in the SHS model.

4.5.1 Simplified Settings

As finding a closed-form of the average age is rather complicated in the general case where each link can transition to SLEEP mode, we focus in the following on a simplified case. In this scenario, we consider N interfering links contending for the channel where only *one* link among the N links, denoted in the sequel by link i, has the ability to transition to SLEEP mode. We study these settings to capture the essence of the extra freedom given to the link and observe its potentials and consequences on the performance of the network. To that extent, in addition to the previous CSMA procedures detailed in Section 4.2, we suppose that link i sleeps for an amount of time upon successful transmission. The sleep duration is assumed to be exponentially distributed with an average of $\frac{1}{w}$. Consequently, the variables to be optimized are[5]: $(R_1, \ldots, R_N, w)$. By taking into account the aforementioned procedures, we define the discrete states set as $\mathbb{Q} = \{0, 1, \ldots, N, 0', 1', \ldots, i'-1, i'+1, \ldots, N'\}$ where $q(t) = k$ if link k has captured the channel and started transmission at time t while $q(t) = 0$ if the channel is idle at time t. The denotation

[5]Note that w may tend to $+\infty$ in the cases where it is optimal for link i to always stay awake. Accordingly, we upper bound w by a large value W_{UB} to make sure that this parameter does not grow to $+\infty$.

"′" refers to the states where link i is in SLEEP mode (e.g., $q(t) = k'$ signifies that link k is transmitting at time t and link i is in SLEEP mode). The Markov chain $q(t)$ is reported in Fig. 4.7.

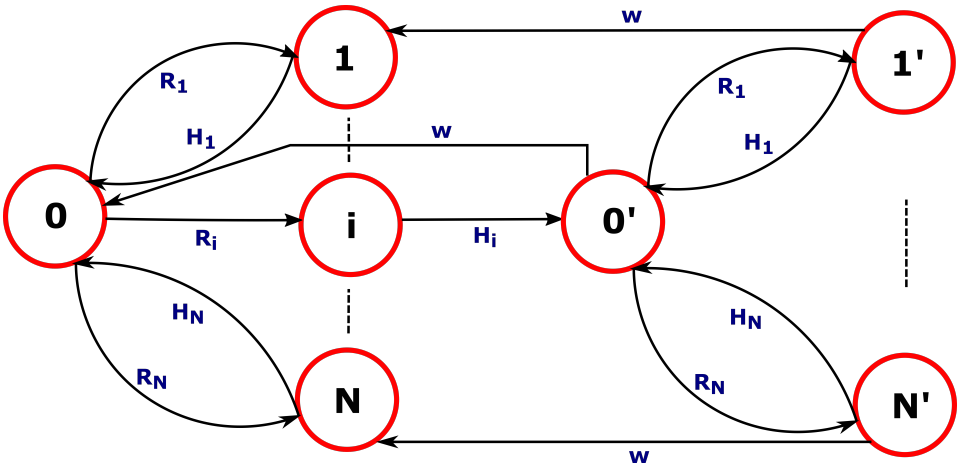


Figure 4.7: Markov chain of the simplified scenario.

To tackle the problem of deriving a closed-form expression of the total average age of the network, we first have to find the stationary distribution of the Markov chain in question. For this purpose, we provide the following proposition.

Proposition 4.2. *The continuous-time Markov chain is irreducible and admits* $\overline{\boldsymbol{\pi}}$ *as stationary distribution where*

$$\overline{\pi}_0 = \frac{1}{C(\boldsymbol{R}) + \frac{R_i}{w + \sum\limits_{\substack{k=1 \\ k\neq i}}^{N} a_k}\left(C(\boldsymbol{R}) - \frac{R_i}{H_i}\right)}, \tag{4.32}$$

$$\overline{\pi}_{0'} = \frac{R_i \overline{\pi}_0}{w + \sum\limits_{\substack{k=1 \\ k\neq i}}^{N} a_k}, \qquad \overline{\pi}_i = \frac{R_i \overline{\pi}_0}{H_i}, \tag{4.33}$$

$$\overline{\pi}_{k'} = \frac{R_k R_i \overline{\pi}_0}{(H_k + w)\left(w + \sum\limits_{\substack{k=1 \\ k\neq i}}^{N} a_k\right)}, \qquad \forall k', \tag{4.34}$$

$$\overline{\pi}_k = \overline{\pi}_0\left(\frac{R_k}{H_k} + \frac{w R_k R_i}{H_k (H_k + w)\left(w + \sum\limits_{\substack{k=1 \\ k\neq i}}^{N} a_k\right)}\right), \qquad \forall k \neq i, \tag{4.35}$$

where

$$C(\boldsymbol{R}) = 1 + \sum_{k=1}^{N} \frac{R_k}{H_k}, \qquad \text{and} \qquad a_k = \frac{w R_k}{H_k + w}, \quad \forall k \neq i. \tag{4.36}$$

Proof. We first start by providing the general balance equations at any state $k' \neq 0'$, which leads to $\overline{\pi}_{k'} = \frac{R_k \overline{\pi}_{0'}}{(H_k+w)}, \ \forall k' \neq 0'$. Next, we formulate the balance equation at state i and we end up with $\overline{\pi}_i = \frac{R_i \overline{\pi}_0}{H_i}$. Afterward, the same is done at state $0'$: $\overline{\pi}_{0'} = \frac{H_i \overline{\pi}_i}{w+\sum\limits_{\substack{k=1\\k\neq i}}^{N} R_k} + \frac{\sum\limits_{\substack{k=1\\k\neq i}}^{N} H_k \overline{\pi}_{k'}}{w+\sum\limits_{\substack{k=1\\k\neq i}}^{N} R_k}$. By replacing our first two results in this equation, we get eq. (4.33) and, consequently, eqs. (4.34). In the last step, we formulate the balance equations at state $k \neq i$: $\overline{\pi}_k = \frac{R_k \overline{\pi}_0}{H_k} + \frac{w \overline{\pi}_{k'}}{H_k} \ \forall k \neq i$. Similarly, by combining these results with those of eqs. (4.34), we end up with eqs. (4.35). Lastly, by taking into account that $\sum_{q\in\mathbb{Q}} \overline{\pi}_q = 1$, $\overline{\pi}_0$ can be calculated, which concludes the proposition. □

With the Markov chain's stationary distribution being characterized, we proceed with finding the total average age of the network. To that end, we distinguish between two perspectives: the perspective of link i and the perspective of links $j \neq i$.

Theorem 4.4. *In the aforementioned system, the average age of link i is $\overline{\Delta}_i(\boldsymbol{R}, w) = \overline{v}_{00} + \overline{v}_{0'0} + \sum\limits_{k=1}^{N} \overline{v}_{k0} + \sum\limits_{\substack{k=1\\k\neq i}}^{N} \overline{v}_{k'0}$ where*

$$\overline{v}_{0'0} = \frac{\overline{\pi}_{0'} + \overline{\pi}_i + \sum\limits_{\substack{k=1\\k\neq i}}^{N} \frac{H_k}{H_k+w} \overline{\pi}_{k'}}{w + \sum\limits_{\substack{k=1\\k\neq i}}^{N} a_k}, \tag{4.37}$$

$$\overline{v}_{k'0} = \frac{\overline{\pi}_{k'}}{H_k + w} + \frac{R_k \overline{v}_{0'0}}{H_k + w}, \quad \forall k', \tag{4.38}$$

$$\overline{v}_{00} = \frac{\overline{\pi}_0 + \sum\limits_{\substack{k=1\\k\neq i}}^{N} \overline{\pi}_k}{R_i} + \frac{w}{R_i}(\overline{v}_{0'0} + \sum_{\substack{k=1\\k\neq i}}^{N} \overline{v}_{k'0}), \tag{4.39}$$

$$\overline{v}_{i0} = \frac{\overline{\pi}_i}{H_i} + \frac{R_i \overline{v}_{00}}{H_i}, \tag{4.40}$$

$$\overline{v}_{k0} = \frac{\overline{\pi}_k}{H_k} + \frac{R_k \overline{v}_{00}}{H_k} + \frac{w \overline{v}_{k'0}}{H_k}, \quad \forall k \neq i, \tag{4.41}$$

and $\overline{\boldsymbol{\pi}}$ is the stationary distribution of the Markov chain reported in Proposition 4.2.

Proof. The proof can be found in Appendix B.4. □

Theorem 4.5. *In the aforementioned system, the average age of link $j \neq i$ is:* $\overline{\Delta}_j(\boldsymbol{R}) = \overline{v}_{00} + \overline{v}_{0'0} + \sum_{k=1}^{N} \overline{v}_{k0} + \sum_{\substack{k=1 \\ k\neq i}}^{N} \overline{v}_{k'0}$ *where*

$$\overline{v}_{00} = \frac{\sum_{\substack{k=0 \\ k\neq i}}^{N} \overline{\pi}_k + \sum_{\substack{k=1 \\ k\neq i}}^{N} \frac{w\overline{\pi}_{k'}}{H_k+w} + \frac{w(\overline{\pi}'_0+\overline{\pi}_i+\sum_{\substack{k=1 \\ k\neq i}}^{N} \frac{H_k\overline{\pi}_{k'}}{H_k+w})}{(w+R_j+\sum_{\substack{k=1 \\ k\neq\{i,j\}}}^{N} a_k)}(1+\sum_{\substack{k=1 \\ k\neq\{i,j\}}}^{N} \frac{R_k}{(H_k+w)})}{R_i + R_j - \frac{wR_i}{(w+R_j+\sum_{\substack{k=1 \\ k\neq\{i,j\}}}^{N} a_k)}(1+\sum_{\substack{k=1 \\ k\neq\{i,j\}}}^{N} \frac{R_k}{(H_k+w)})}, \tag{4.42}$$

$$\overline{v}_{0'0} = \frac{\overline{\pi}_{0'} + \overline{\pi}_i + \sum_{\substack{k=1 \\ k\neq i}}^{N} \frac{H_k\overline{\pi}_{k'}}{H_k+w} + R_i\overline{v}_{00}}{w + R_j + \sum_{\substack{k=1 \\ k\neq\{i,j\}}}^{N} a_k}, \tag{4.43}$$

$$\overline{v}_{k'0} = \frac{\overline{\pi}_{k'}}{H_k+w} + \frac{R_k\overline{v}_{0'0}}{H_k+w}, \quad \forall k', \tag{4.44}$$

$$\overline{v}_{i0} = \frac{\overline{\pi}_i}{H_i} + \frac{R_i\overline{v}_{00}}{H_i}, \tag{4.45}$$

$$\overline{v}_{k0} = \frac{\overline{\pi}_k}{H_k} + \frac{R_k\overline{v}_{00}}{H_k} + \frac{w\overline{v}_{k'0}}{H_k}, \quad \forall k \neq i, \tag{4.46}$$

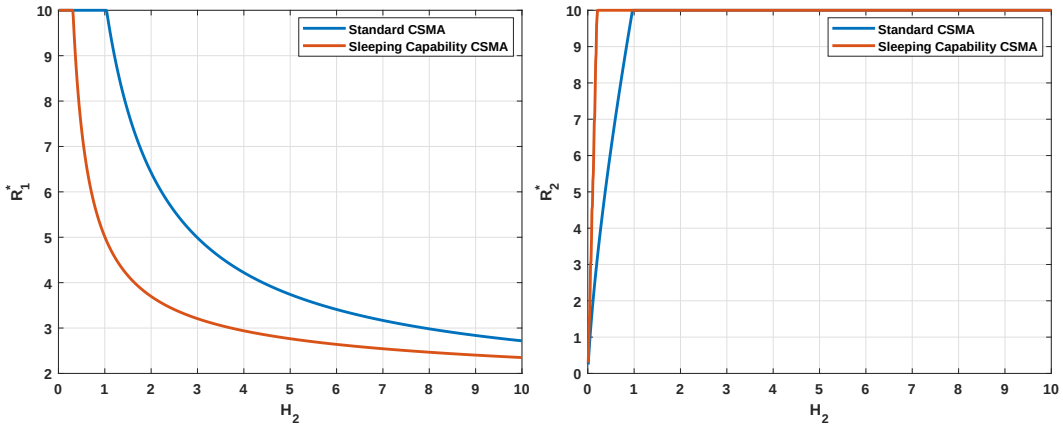
and $\overline{\boldsymbol{\pi}}$ is the stationary distribution of the Markov chain reported in Proposition 4.2.

Proof. The proof can be found in Appendix B.5. □

To evaluate the overall network's performance, we note that the total average age is $\overline{\Delta} = \overline{\Delta}_i + \sum_{\substack{j=1 \\ j\neq i}}^{N} \overline{\Delta}_j$. With the closed-form expression being found, we investigate the benefits of the extra freedom given to link i. For this purpose, we consider a two links scenario where $H_1 = 1$, $R_{UB} = 10$, $W_{UB} = 100$ and we iterate over a range of H_2. Link 2 has the freedom to go to SLEEP mode. We compare the optimal back-off rates (R_1^*, R_2^*) of the proposed scheme with those of the optimized CSMA scheme of Section 4.4 in Fig. 4.8a and Fig. 4.8b. Moreover, the optimal waking-up rate w^* is plotted along with the age gain provided by the scheme compared to the optimized standard CSMA in Fig. 4.9a and Fig. 4.9b respectively. We summarize our findings in the following.

The first conclusion we can draw from these results is that R_1^*, in our proposed scheme, is always smaller or equal to that in standard CSMA, as seen in Fig. 4.8a.

This is because link 2 spends an amount of time in SLEEP mode and reduces the burden on the channel. Therefore, link 1 does not need to be as aggressive on the channel. On the counterpart, we can see in Fig. 4.8b that R_2^* is always higher in our proposed scheme. This is because link 2 needs to compensate for its time spent in SLEEP mode, and therefore it is more aggressive on the channel when it is awake.



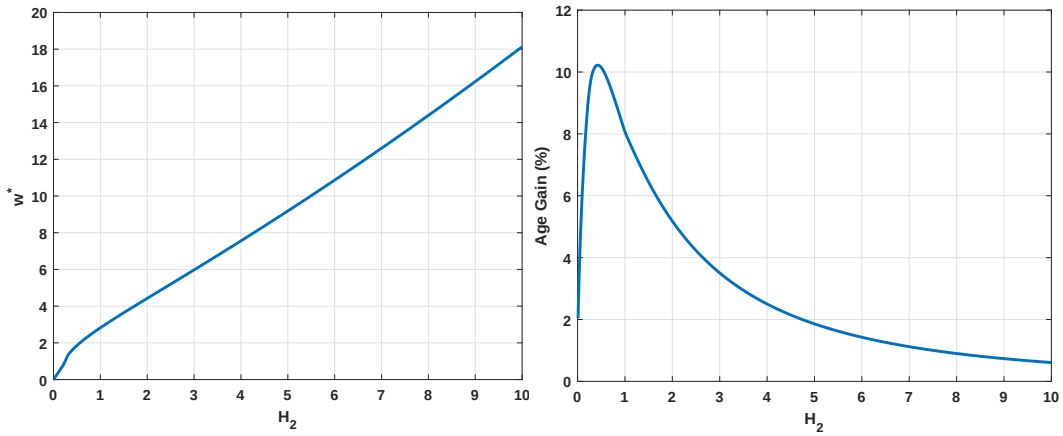
(a) Optimal back-off rate R_1^*.

(b) Optimal back-off rate R_2^*.

Figure 4.8: Comparison between the two schemes in terms of R_1^* and R_2^*.

The second conclusion we can draw is that w^* is always increasing with H_2 as seen in Fig. 4.9a. To understand this trend, we recall that, as shown in the previous section, links with fast service rates should be given priority to access the channel in CSMA environments. Accordingly, we can see that as H_2 increases, the optimal sleeping rate grows. This lets link 2 quickly transition back to AWAKE mode to capture the opportunity presented by the fast service rate H_2. Bearing in mind the two above conclusions, we can now understand the trend of the gain in total average age reported in Fig. 4.9b. For extremely low H_2, link 1 should always get priority on the channel as it is the link with the faster service rate. Therefore, R_1^* is equal to R_{UB} and R_2 is extremely low in both schemes. Consequently, the age gain is minor in this regime. As H_2 increases, the benefit of adopting a sleeping mechanism by link 2 to reduce the burden on the channel starts to show up. This gain reaches around 10% for $H_2 = 0.5$. When H_2 increases beyond this point, the age gain decreases, and the two schemes' performance coincides. This is because w^* is increasing, and therefore link 2 barely stays in SLEEP mode upon successful transmission.

The above results provide us with insights that capture our proposed scheme's essence and show the benefits of the extra degree of freedom given to link 2. Next, we define the age gain as the ratio of the average age difference between the two policies divided by the optimal standard CSMA scheme's average age. In other words, it represents the amount of age reduction that we end up with by adopting the improved CSMA scheme. As one can see in Fig. 4.9b, the reduction in average age, which we will call the age gain, can be as high as 10%, and the potentials of this scheme can be witnessed. Since even higher gains can be achieved when providing



(a) Optimal waking-up rate w^*.

(b) Gain in average age.

Figure 4.9: Comparison between the two schemes in terms of w^* and average age.

the same freedom to all links instead of just one, we tackle this general case in the next subsection.

4.5.2 General Scenario

This section investigates the general case where all links can transition to SLEEP mode upon successful transmission. In addition to the CSMA procedures detailed in Section 4.2, we suppose that each link k sleeps for an amount of time upon successful transmission. The sleep duration is assumed to be exponentially distributed with an average of $\frac{1}{w_k}$. Consequently, the variables to be optimized are[6]: $(R_1, \ldots, R_N, w_1, \ldots, w_N)$. We proceed with our SHS analysis, and we define the discrete process $q(t)$ as a $2D$ Markov chain where each state (a^j, x^i) is composed of:

- a^j: we refer to it as the *configuration state*, and it is an N-tuplet of binary variables a_k^j that indicates if link k is awake (binary value 1) or asleep (binary value 0). In fact, we have 2^N possible configuration states for the links.

- x^i: for each configuration state a^j, we denote by x^i the *transmission state*, which is an N-tuplet of binary variables x_k^i that indicates if link k is transmitting or not. Clearly, x_k^i cannot be equal to 1 unless link k is actually awake (i.e., $a_k^j = 1$). Moreover, due to interference, x^i can only take two values: 1) Null vector $\mathbf{0}_N$, and 2) $x^i = \boldsymbol{e}_k$ where $\boldsymbol{e}_k$ represents the canonical vector in $\mathbb{R}^N$. The canonical vector is defined as $(\boldsymbol{e}_k)_{l:1\leq l\leq N} = (\delta_{kl})_{1\leq l\leq N}$ where δ_{kl} is the Kronecker delta function. In this context, we have that the total number of transmission states x^i for a fixed configuration state a^j is $|I_j| = 1 + \sum_{k=1}^{N} a_k^j$.

[6]For the same reason as the previous section, we upper bound w_k, $\forall k$ by a large real value W_{UB}.

It is straightforward that this $2D$ Markov chain is ergodic by construction for $\boldsymbol{R} > \boldsymbol{0}, \boldsymbol{w} > \boldsymbol{0}$, and therefore admits a unique stationary distribution. Consequently, we define the compact convex set $\mathcal{X}$ such that each element $\boldsymbol{x} = [\boldsymbol{R}, \boldsymbol{w}] \in \mathcal{X}$ verifies the following conditions: $\zeta \leq R_k \leq R_{UB}, \ \forall k$ and $\zeta \leq w_k \leq W_{UB}, \ \forall k$, where $\zeta > 0$ is an arbitrarily small constant[7].

The next step in our SHS analysis consists of finding the stationary distribution $\overline{\boldsymbol{\pi}}$ of the Markov chain in question. Therefore, we formulate the general balance equations for each state (a^j, x^i) in Table 4.2. By taking these equations into account, and knowing that $\sum_{j=1}^{2^N} \sum_{i=1}^{|I_j|} \overline{\pi}(a^j, x^i) = 1$, we can deduce the matrix $\boldsymbol{A} \in \mathbb{R}^{S \times S}$ and vector $\boldsymbol{d} \in \mathbb{R}^S$ where S is the number of states of the Markov chain such that $\boldsymbol{A}\overline{\boldsymbol{\pi}} = \boldsymbol{d}$. In this framework, we have that $\overline{\boldsymbol{\pi}} = [\overline{\pi}(\boldsymbol{0}_N, \boldsymbol{0}_N), \ldots, \overline{\pi}(\boldsymbol{1}_N, \boldsymbol{e}_N)]^T$ and $\boldsymbol{d} = [0, \ldots, 0, 1]^T$. As the uniqueness of $\overline{\boldsymbol{\pi}}$ is established due to ergodicity of the underlying Markov chain, we can assert the existence of the inverse matrix $\boldsymbol{A}^{-1}$ such that $\overline{\boldsymbol{\pi}} = \boldsymbol{A}^{-1}\boldsymbol{d}$.

We continue with the second step of our SHS analysis and we provide in Table 4.2 the detailed application of eq. (3.6) from the perspective of a link of interest m. Based on them, we can find the matrices $\boldsymbol{E}_m \in \mathbb{R}^{2S \times 2S}$ and $\boldsymbol{F}_m \in \mathbb{R}^{2S \times S}$ such that $\boldsymbol{E_m v}^m = \boldsymbol{F}_m \overline{\boldsymbol{\pi}}$ where $\overline{\boldsymbol{v}}^m = [\overline{v}^m(\boldsymbol{0}_N, \boldsymbol{0}_N)_0, \overline{v}^m(\boldsymbol{0}_N, \boldsymbol{0}_N)_1, \ldots, \overline{v}^m(\boldsymbol{1}_N, \boldsymbol{e}_N)_1]^T$. As explained in Theorem 3.1, due to the ergodicity of the Markov chain, we have that $\boldsymbol{E}_m^{-1}$ exists, and therefore $\overline{\boldsymbol{v}}^m = \boldsymbol{E}_m^{-1}\boldsymbol{F}_m\overline{\boldsymbol{\pi}} = \boldsymbol{E}_m^{-1}\boldsymbol{F}_m\boldsymbol{A}^{-1}\boldsymbol{d}$. As the average age[8] of link m is nothing but the sum of the 0 indices components of $\overline{\boldsymbol{v}}^m$ (we recall the results of Theorem 3.1), we extract these components through multiplication by the row vector $\boldsymbol{c} \in \mathbb{R}^{2S}$ where the vector $\boldsymbol{c}$ is composed of binary elements 1 and 0 corresponding to the 0 and 1 indices of $\overline{\boldsymbol{v}}^m$ respectively. Consequently, our optimization problem becomes:

$$\underset{\boldsymbol{x} \in \mathcal{X}}{\text{minimize}} \quad \overline{\Delta}(\boldsymbol{x}) = \sum_{m=1}^{N} \boldsymbol{c}\boldsymbol{E}_m^{-1}\boldsymbol{F}_m\boldsymbol{A}^{-1}\boldsymbol{d}. \tag{4.50}$$

Since finding a closed-form of the objective function in (4.50) is rather complicated, we seek to circumvent this difficulty by employing a sequential convex approximation approach. To that extent, the proposed SCA approach can be summarized in the following:

$$\hat{\boldsymbol{x}}[n] = \underset{\boldsymbol{x} \in \mathcal{X}}{\operatorname{argmin}} \ \overline{\Upsilon}(\boldsymbol{x}, \overline{\boldsymbol{x}}[n]), \quad n = 1, 2 \ldots, \tag{4.51}$$

where

$$\overline{\Upsilon}(\boldsymbol{x}, \overline{\boldsymbol{x}}[n]) = \overline{\Delta}(\overline{\boldsymbol{x}}[n]) + \nabla\overline{\Delta}(\overline{\boldsymbol{x}}[n])^T(\boldsymbol{x} - \overline{\boldsymbol{x}}[n]) + \frac{1}{2\alpha_n}||\boldsymbol{x} - \overline{\boldsymbol{x}}[n]||_2^2. \tag{4.52}$$

The term $\frac{1}{2\alpha_n}||\boldsymbol{x} - \overline{\boldsymbol{x}}[n]||_2^2$ is used to regularize the approximation and keep the points close enough so that the model is accurate. Conditions on α_n to ensure the convergence of the approach will be presented in Proposition 4.3. As the problem in (4.51)

[7] Defining the compact convex set $\mathcal{X}$ will be vital to the convergence proof in Proposition 4.3.

[8] Our interest lays in the case where the underlying SHS is stable, i.e., $\overline{v}^m, \ \forall m$ is non-negative [64].

is convex, it can be solved using any standard convex solvers such as CVX [81]. After finding the solution of (4.51), at each iteration, we set $\overline{\boldsymbol{x}}[n+1] = \hat{\boldsymbol{x}}[n]$. The next step consists of finding the expression of the gradient of the average age. To do so, we observe that $\frac{\partial \overline{\Delta}(\boldsymbol{x})}{\partial x_i} = \sum_{m=1}^{N} \boldsymbol{c} \frac{\partial \boldsymbol{E}_m^{-1}(\boldsymbol{x})}{\partial x_i} \boldsymbol{F}_m \boldsymbol{A}^{-1}(\boldsymbol{x}) \boldsymbol{d} + \boldsymbol{c} \boldsymbol{E}_m^{-1}(\boldsymbol{x}) \boldsymbol{F}_m \frac{\partial \boldsymbol{A}^{-1}(\boldsymbol{x})}{\partial x_i} \boldsymbol{d}$. Using the following identity $\frac{\partial \boldsymbol{K}^{-1}}{\partial x_i} = -\boldsymbol{K}^{-1} \frac{\partial \boldsymbol{K}}{\partial x_i} \boldsymbol{K}^{-1}$, we can conclude that $\frac{\partial \overline{\Delta}(\boldsymbol{x})}{\partial x_i} = -\sum_{m=1}^{N} \boldsymbol{c} \boldsymbol{E}_m^{-1} \frac{\partial \boldsymbol{E}_m}{\partial x_i} \boldsymbol{E}_m^{-1} \boldsymbol{F}_m \boldsymbol{A}^{-1} \boldsymbol{d} - \sum_{m=1}^{N} \boldsymbol{c} \boldsymbol{E}_m^{-1} \boldsymbol{F}_m \boldsymbol{A}^{-1} \frac{\partial \boldsymbol{A}}{\partial x_i} \boldsymbol{A}^{-1} \boldsymbol{d}$. Based on the above equation, the gradient vector $\nabla \overline{\Delta}(\boldsymbol{x})$ can be found. We summarize our approach in Algorithm 1.

Remark 4.6. *It is worth noting that the resulting matrices $\boldsymbol{E}_m, \boldsymbol{F}_m, \ \forall m$ and $\boldsymbol{A}$ from equations (4.47)-(4.49) are sparse by nature. Consequently, the same can be said about the matrices $\frac{\partial \boldsymbol{E}_m}{\partial x_i}$ and $\frac{\partial \boldsymbol{A}}{\partial x_i}$ whose entries have the values $0, 1$ and -1. This is very appealing in practice as it allows faster and more efficient implementations of the required matrix operations (e.g., multiplication [82], inversion [83]).*

Algorithm 1 Proposed SCA approach

1: **Input** Stopping criterion ϵ and two feasible points $\overline{\boldsymbol{x}}[1], \hat{\boldsymbol{x}}[1] \in \mathcal{X}$
2: **Initialize** Set $n = 1$
3: **Iterate**
4: $\overline{\boldsymbol{x}}[n+1] := \hat{\boldsymbol{x}}[n]$
5: $n := n + 1$
6: Solve the convex problem in (4.51) to find $\hat{\boldsymbol{x}}[n]$
7: **Until** $||\overline{\Upsilon}(\hat{\boldsymbol{x}}[n], \overline{\boldsymbol{x}}[n]) - \overline{\Upsilon}(\hat{\boldsymbol{x}}[n-1], \overline{\boldsymbol{x}}[n-1])|| < \epsilon$
8: Output $\overline{\boldsymbol{x}}[n]$

In the sequel, we provide a convergence analysis of the algorithm presented above. To proceed in that direction, we first lay out the following definition.

Definition 4.1 (Stationary points of a function)**.** *Let $f : \mathcal{D} \rightarrow \mathbb{R}$ be a function where $\mathcal{D} \subseteq \mathbb{R}^n$ is a convex set. A point $\boldsymbol{x}^* \in \mathcal{D}$ is a stationary point of $f(.)$ if $\nabla_{\vec{\boldsymbol{d}}} f(\boldsymbol{x}^*) \geq 0$ for all $\vec{\boldsymbol{d}} \in \mathcal{D}$ such that $\boldsymbol{x}^* + \vec{\boldsymbol{d}} \in \mathcal{D}$.*

Equipped with the above definition, we present the following convergence results.

Proposition 4.3. *The sequence $\left\{\overline{\Upsilon}(\hat{\boldsymbol{x}}[n], \overline{\boldsymbol{x}}[n])\right\}_{n=1}^{+\infty}$ is convergent for $\alpha_n \geq \frac{L}{2}, n \in \mathbb{N}$ where L is the Lipschitz constant of the function $\nabla \overline{\Delta}(\boldsymbol{x})$. Moreover, the limit point of the sequence $\left\{\overline{\boldsymbol{x}}[n]\right\}_{n=1}^{+\infty}$ generated by the SCA procedure (4.51) is a stationary point of the problem in (4.50).*

Proof. The proof can be found in Appendix B.6 of the supplementary material. □

Remark 4.7. *We point out that our proposed SCA approach is general and holds for any minimization, on a compact convex set, of the average age of a stable system modeled through the same SHS tools.*

4.5.3 SCA Approach Implementation

In the following, we investigate the performance advantage of our proposed SCA scheme compared to the optimal standard CSMA. To that extent, we consider a two links scenario where $H_1 = 1$, $R_{UB} = 15$, $W_{UB} = 100$, and $\zeta = 10^{-6}$. Contrary to the previous section, we provide both links the freedom to go to SLEEP mode upon successful transmission. The initial point $\boldsymbol{x}[1]$ is set to be any random point belonging to $\mathcal{X}$. We plot the optimal performance parameters $(R_1^*, R_2^*, w_1^*, w_2^*)$ along with the total gain in total average age in function of H_2, as depicted in the following.

In Fig. 4.10, one can see that the waking-up rates' curves intersect for $H_2 = H_1 = 1$. Before this point, link 2 sleeps more than link 1 while the opposite happens on the other side of the point. The reason behind this has been extensively explained in previous sections: faster service rate links provide an opportunity that needs to be captured. Therefore, links with slow service rate sleep more to reduce the burden on the channel to allow easier capture of the channel by faster service rate links. Moreover, we can see that links with a higher service rate always end up being more aggressive on the channel for the same reason.

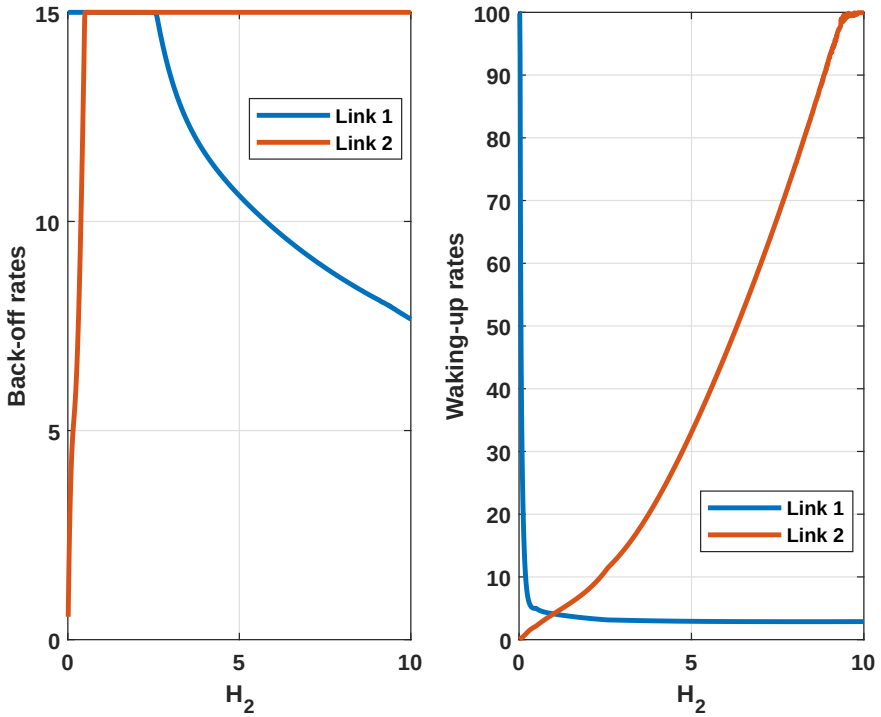


Figure 4.10: The optimal performance parameters.

As for the results in Fig. 4.11, we can see that the proposed scheme exhibits a continuous gain compared to the standard CSMA. The gain can be as high as 16% in the two links scenario. This highlights the potentials of our newly proposed scheme compared to the optimized standard CSMA.

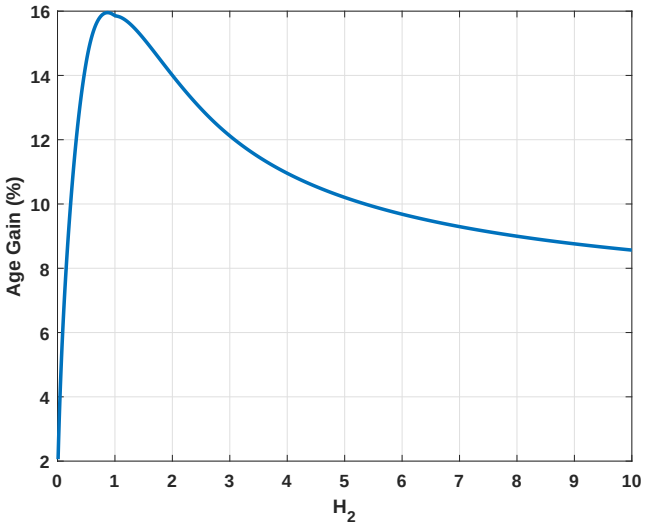


Figure 4.11: The gain in total average age.

4.6 Conclusion

In this chapter, we have investigated the AoI in a CSMA environment where N links contend for the channel. Using SHS tools, we derived a closed-form expression of the average age in two distinct scenarios: 1) links generate packets at will, and 2) packets arrive stochastically to each link. An optimization problem to minimize the total average age was then formulated. An equivalent convex formulation was presented by investigating the derived problem, which makes finding the optimal back-off time of each link a simple task. Interestingly, it was shown that the minimum average age is achieved for the same back-off duration in both the sampling and the stochastic arrivals scenarios. Numerical implementations of our proposed solution in an IEEE 802.11 network were provided, and its performance was highlighted in function of the nodes' density. Also, motivated by further improving the performance, we proposed a new modified CSMA scheme that outperforms the optimized standard CSMA. Simulations results were then laid out to highlight its performance in comparison to that of the optimized standard CSMA.

$$\overline{\pi}(a^j,x^i)\Big(\sum_{k=1}^{N}\big[(1-a_k^j)w_k+a_k^jx_k^iH_k+(1-\sum_{l=1}^{N}a_l^jx_l^i)a_k^jR_k\big]\Big)=$$
$$\sum_{k=1}^{N}\big[a_k^j(1-x_k^i)\overline{\pi}(a^j-\boldsymbol{e}_k,x^i)w_k+(1-\sum_{l=1}^{N}a_l^jx_l^i)(1-a_k^j)H_k\overline{\pi}(a^j+\boldsymbol{e}_k,x^i+\boldsymbol{e}_k)$$
$$+(\sum_{l=1}^{N}a_l^jx_l^i)x_k^jR_k\overline{\pi}(a^j,x^i-\boldsymbol{e}_k)\big],\quad\forall(a^j,x^i),\quad(4.47)$$

$$\overline{v}^m(a^j,x^i)_0\Big(\sum_{k=1}^{N}\big[(1-a_k^j)w_k+a_k^jx_k^iH_k+(1-\sum_{l=1}^{N}a_l^jx_l^i)a_k^jR_k\big]\Big)=\overline{\pi}(a^j,x^i)$$
$$+\sum_{k=1}^{N}\big[a_k^j(1-x_k^i)w_k\overline{v}(a^j-\boldsymbol{e}_k,x^i)_0+(\sum_{l=1}^{N}a_l^jx_l^i)x_k^jR_k\overline{v}(a^j,x^i-\boldsymbol{e}_k)_0\big]$$
$$+\sum_{\substack{k=1\\k\neq m}}^{N}\big[(1-\sum_{l=1}^{N}a_l^jx_l^i)(1-a_k^j)H_k\overline{v}(a^j+\boldsymbol{e}_k,x^i+\boldsymbol{e}_k)_0\big]$$
$$+(1-\sum_{l=1}^{N}a_l^jx_l^i)(1-a_m^j)H_m\overline{v}(a^j+\boldsymbol{e}_k,x^i+\boldsymbol{e}_k)_1,\quad\forall(a^j,x^i),\forall m\quad(4.48)$$

$$\overline{v}^m(a^j,x^i)_1\Big(\sum_{k=1}^{N}\big[(1-a_k^j)w_k+a_k^jx_k^iH_k+(1-\sum_{l=1}^{N}a_l^jx_l^i)a_k^jR_k\big]\Big)=a_m^jx_m\overline{\pi}(a^j,x^i)$$
$$+a_m^jx_m\sum_{k=1}^{N}a_k^j(1-x_k^i)w_k\overline{v}(a^j-\boldsymbol{e}_k,x^i)_1,\quad\forall(a^j,x^i),\forall m.\quad(4.49)$$

Table 4.2: The general balance equations and the resulting equations from the SHS approach.

Part II

Age of Incorrect Information

5 | Age of Incorrect Information: Analysis and Optimization

5.1 Overview

This chapter introduces a new performance metric in the framework of status updates systems, referred to as the age of incorrect information. As previously explained in the introduction, the ultimate goal in a status update system is to have the best real-time remote estimation of the process of interest at the monitor side. The shortcomings of the AoI were put into perspective in various real-time estimation problems. For instance, when X_t is a Wiener process and sampling times are allowed to depend on X_t, the MMSE policy in a delay channel is not equivalent to a minimization of an age-based metric [40]. This stems from the fact that the AoI, by definition, does not capture well the information content of the transmitted packets nor the current estimate at the monitor. Another critical question that arises is the following: should the minimization of prediction error or mean squared error always be regarded as the definitive goal of the remote estimation scenario? To argue that this should not always be the case, we shed light on one of the shortcomings of these conventional error measures. The primary issue with these error functions is that they do not increasingly penalize the monitor for wrongfully estimating the process of interest. Namely, the same penalty is paid for being in an erroneous state no matter how long the monitor has been in it. To that extent, a monitor wrongfully thinking that a machine is at a reasonable temperature suffers from the same penalty no matter how long it has been overheating. This suggests that a more general framework should be introduced to address the shortcomings of these error measures. In this chapter, we pave the way for such a framework. We do so by introducing a new performance metric that deals with the above shortcomings of both the AoI and the error functions. To that end, we summarize in the following the key contributions of this chapter:

- We first go into more depth on highlighting the shortcomings of the AoI and the error performance metrics in the case of remote process estimation. To deal with these shortcomings, we propose a new performance measure, which we call the age of incorrect information. This measure neatly extends the no-

tion of fresh updates to that of fresh “informative" updates. The word informative refers to updates that bring new and correct information to the monitor side. This new measure also captures the deteriorating effect the wrong information can have with time on the system.

- Afterward, we focus on the case where a transmitter-receiver pair communicates over an unreliable channel. The transmitter sends status updates about an N states Markovian information source with the goal of the receiver being to estimate it accurately. In this scenario, we aim to find the optimal transmission policy that minimizes the average proposed metric. By casting this problem as an MDP, we show that when no constraints on the power are imposed, an “always update" policy can minimize the average age, the prediction error, and the average AoII.

- Following that, we tackle the more realistic case where each transmission incurs a cost, and the transmitter has a power budget that cannot be surpassed. We cast our problem into a Constrained Markov Decision Process (CMDP) that is known to be challenging to solve. To circumvent this difficulty, we provide a Lagrange approach that transforms the CMDP to an unconstrained MDP. The Lagrangian optimization problem is then thoroughly studied, and structural results on its optimal policy are provided.

- Subsequently, we provide a rigorous mathematical proof to show that the optimal operating point of the CMDP is achieved by a mixture of two deterministic Lagrange policies. Similar results were established in the literature for the AoI optimization framework [23]. However, due to the inherent properties of the proposed AoII metric, the standard approach adopted in [23] cannot be followed. Accordingly, we proceed in a different direction to establish the required results, as will be seen in later sections of the chapter. Equipped with these results, we provide an algorithm that finds the AoII-optimal policy under the power constraint in logarithmic complexity.

- Lastly, we provide numerical implementations of our transmission policy. These implementations showcase interesting insights on the differences between the AoI, the standard error penalty, and the AoII frameworks.

The rest of the chapter is organized as follows: Section 5.2 is dedicated to the motivation of the newly proposed framework. The system model, along with the dynamics of the proposed metric are presented in Section 5.3. Section 5.4 provides the MDP description of the problem along with its analysis in the unconstrained power scenario. In Section 5.5, we thoroughly analyze the constrained scenario and propose an optimal approach to solve it. Numerical results that corroborate the theoretical findings are laid out in Section 5.6, and conclusions are presented in Section 5.7.

5.2 Proposed Metric

To put our work into perspective, we focus in this section on a particular scenario where a transmitter-receiver pair communicates. More specifically, the transmitter observes a process $X(t)$ and informs the receiver (monitor) about it by sending status updates over the network. Based on the last received update, the monitor constructs an estimate of the process, denoted by $\hat{X}(t)$. Time is considered to be discrete and normalized to the time slot duration. For simplicity, we suppose in this section that the process in question can only have two values $\{1, 2\}$, as depicted in Fig 5.1. At each time slot, the probability of remaining in the same state is p_R, while the probability of transitioning to another state is p_t. The transmitter decides when to inform the monitor about the process $X(t)$ by adopting a transmission policy that aims to minimize the average of a particular penalty function.

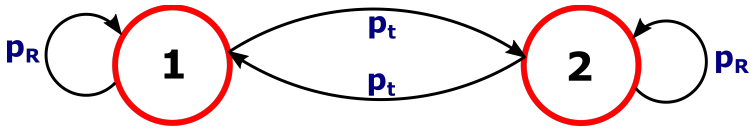


Figure 5.1: Example of a two states information source.

First, let us consider the age penalty function to examine its shortcomings more closely. To that extent, we define the age as

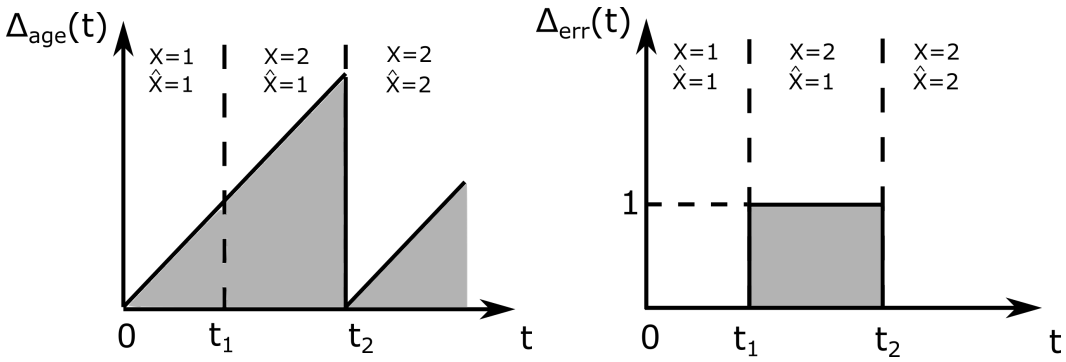
$$\Delta_{\text{age}}(t) = t - U(t), \tag{5.1}$$

where $U(t)$ is the time-stamp of the last successfully received packet by the monitor. Based on this definition, we can observe that the age captures the information *time lag* at the monitor, in an attempt to achieve timely updates. As seen in (5.1), the age always increases as time progresses regardless of the current information at the monitor, which makes it fall short in numerous applications. To see this, let us observe the trend of $\Delta_{\text{age}}(t)$ in the time interval $[0, t_1]$ of Fig. 5.2a. In this interval, the monitor has perfect knowledge of the process of interest $X(t)$. Therefore, any new update received in this interval will not change the information currently available at the monitor. Regardless of that, we can clearly see that the age penalty keeps growing with time, i.e., a penalty is being paid for not being updated on the information process. However, the monitor currently has perfect knowledge of the process. This above observation puts into perspective the shortcoming of the age penalty function and let us emphasize on the fact that any relevant metric for the remote estimation of a process has to capture more meaningfully its information content and the current knowledge at the receiver.

Another widely used penalty function is the error penalty:

$$\Delta_{\text{err}}(t) = \mathbb{1}\{\hat{X}(t) \neq X(t)\}, \tag{5.2}$$

In fact, minimizing the average of the function in (5.2), is equivalent to the minimization of the prediction error $\Pr(\hat{X}(t) \neq X(t))$. The key shortcoming of this



(a) Age penalty function.

(b) Error penalty function.

Figure 5.2: Illustrations of the age and error penalty functions.

error penalty function is its failure to capture the following phenomena that arises in numerous applications: staying in an erroneous state should have an increasing penalty effect. In fact, the function in (5.2) treats all instances of error equally, no matter how long the time elapsed since their start is. In other words, the penalty of being in an erroneous state after 1 time slot or 100 time slots is the same value 1. Because of this observation, we can see that the long-time average error penalty due to a burst error is the same as the one resulting from several isolated errors of the same duration. However, this is not always the case. There is a vast amount of applications where the penalty grows the longer the monitor has incorrect information. For example, let us suppose that $X(t) = 1$ refers to the case where a machine is at a normal temperature at time t and $X(t) = 2$ is the case where the machine is overheating. This information has to be transferred to a monitor that can, consequently, react to the machine's state. By considering the time interval $[t_1, t_2]$ of Fig. 5.2b, we can see that no matter how long the duration of the interval $\Delta t = t_2 - t_1$, the same penalty $\Delta_{err}(t) = 1$ is kept. However, as it is well-known, the repercussions of keeping a machine overheated become more severe as time goes on. Therefore, this should be reflected in the adopted penalty function and should be considered one of its key design features. It is worth mentioning that the list of such real-life applications where the level of dissatisfaction grows as time progresses is vast. We report a few examples in the following:

- A real-time video stream in which packets are sent through a channel, and where losses can occur due to, for example, an inaccurate channel estimate. Similarly to the previous case, adopting the AoI as a performance metric will fall short since a penalty is constantly paid even if the current channel estimate is accurate. On another note, if any standard error penalty function is adopted, the effect of burst errors on the performance is not captured. However, in this application, it is well-known that a burst packet losses lead to more distortion of the video compared to an equal number of isolated losses.

- An actuator that can tolerate inaccurate actions for a brief amount of time;

however, when these actions are done for a long duration, substantial performance penalties are to be paid.

- The relay of fire outbreaks in environmental monitoring applications where any relay failure cause more severe repercussions the longer it lasts.

Motivated by all this, we aim to propose in our paper a new metric that elegantly combines the following two characteristics of the age and the error penalty functions:

1. The proposed metric captures the information content of the updates and the monitor's current knowledge as done by the error penalty function in (5.2).
2. The proposed metric captures the increasing dissatisfaction with time that is offered by the age penalty.

Based on this, the general metric that we are about to introduce can be thought to capture the notion of ***fresh informative*** updates. The word informative in this context refers to updates that bring ***new*** information to the monitor side. In other words, when the monitor already has perfect knowledge about the process in question, we should not pay any penalty. However, as the state of the process changes and the monitor becomes in an erroneous state, an update from the transmitter becomes informative. Because we need this update to arrive as ***fresh*** as possible, we let the penalty grows with time as long as we are in an erroneous state. To that extent, our proposed metric, which we will call the age of incorrect information, can be written as follows:

$$\Delta_{\text{AoII}}(t) = f(t) \times g(X(t), \hat{X}(t)), \tag{5.3}$$

where $f(t)$ is non-decreasing time penalty function, paid for being unaware of the process's correct status for a certain amount of time. On the other hand, $g(X(t), \hat{X}(t))$ is an information penalty function that reflects the difference between the current estimate at the monitor and the actual state of the process. There exists a wide variety of choices for f and g that can be picked. We list below some of these examples, starting with g and following it by f.

- The indicator error function:

$$g_{\text{ind}}(X(t), \hat{X}(t)) = \mathbb{1}_{\{X(t) \neq \hat{X}(t)\}}. \tag{5.4}$$

  This information penalty function can be adopted when any mismatch between $X(t)$ and $\hat{X}(t)$ penalizes the system in the same fashion.

- The squared error function:

$$g_{\text{sq}}(X(t), \hat{X}(t)) = (X(t) - \hat{X}(t))^2. \tag{5.5}$$

  Unlike $g_{\text{ind}}(X(t), \hat{X}(t))$, this information penalty function penalizes more the system the larger the difference between $X(t)$ and $\hat{X}(t)$ is.

- The threshold error function:

$$g_{\text{threshold}}(X(t), \hat{X}(t)) = \mathbb{1}_{|X(t)-\hat{X}(t)|\geq c}, \tag{5.6}$$

where $c > 0$ is a predefined threshold. This information penalty function can be used when the system can tolerate small mismatches between $X(t)$ and $\hat{X}(t)$. However, when the mismatch between the two is high, a penalty is paid.

Next, we provide examples of the time-dissatisfaction function f. To do so, we first define $V(t)$ as the last time instant where $g(X(t), \hat{X}(t))$ was equal to 0. In other words, $V(t)$ is the last time instant where the monitor had zero information penalty, i.e. when the monitor had accurate information about the source. By leveraging this notion, we present the following examples of f.

- The linear time-dissatisfaction function:

$$f_{\text{linear}}(t) = t - V(t). \tag{5.7}$$

- The exponential time-dissatisfaction function:

$$f_{\text{exponential}}(t) = \exp(a(t - V(t))), \tag{5.8}$$

where $a > 0$ is a positive constant. This time-dissatisfaction function can be used when the system is extremely vulnerable to wrong information, and the need for fresh correct information snowballs with time.

- The time-threshold dissatisfaction function:

$$f_{\text{threshold}}(t) = \mathbb{1}_{\{t-V(t)\geq d\}}, \tag{5.9}$$

where $\mathbb{1}_{\{.\}}$ is the indicator function, and $d > 0$ is a fixed time threshold that should not be violated. This time-dissatisfaction function can be adopted when the system's performance starts deteriorating due to wrong information beyond a particular duration of time d.

For simplicity, we focus in the sequel on the case where $f(t) = f_{\text{linear}}$ and $g(X(t), \hat{X}(t)) = g_{\text{ind}}$. Specifically, we have:

$$\Delta_{\text{prop}}(t) = f(t) \times g(X(t), \hat{X}(t)) = (t - V(t))\mathbb{1}\{\hat{X}(t) \neq X(t)\}. \tag{5.10}$$

A sketch of this function is given in Fig. 5.3 where we can see how the penalty increases with time in the interval $[t_1, t_2]$ to reflect the increasing dissatisfaction of being in an erroneous state. This metric will be the basis of our analysis in the upcoming sections, where we aim to minimize its average in a general scenario of interest. With that in mind, we stress the fact that our proposed metric is far more general and is not limited to this choice of f and g.

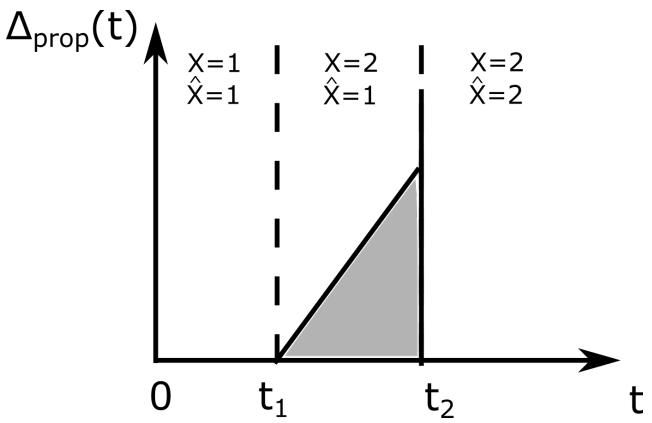


Figure 5.3: Illustration of the proposed penalty function.

5.3 System Overview

5.3.1 System Model

In this chapter, we consider a transmitter-receiver pair where the transmitter sends status updates about the process of interest to the receiver side over an unreliable channel. Time is considered to be slotted and normalized to the slot duration (i.e., the slot duration is taken as 1). The information process of interest is an N state discrete-time Markov chain $\big(X(t)\big)_{t\in\mathbb{N}}$ depicted in Fig. 5.4. To that extent, we define the probability of transitions as

$$\Pr(X(t+1)=j|X(t)=i)=\begin{cases}p_R, & j=i,\\ p_t, & j\neq i.\end{cases} \tag{5.11}$$

Since the process in question can have one of N different possible values, the following always holds:

$$p_R+(N-1)p_t=1. \tag{5.12}$$

As for the unreliable channel model, we suppose that the channel realizations are Independent and Identically Distributed (iid) over the time slots and follow a Bernoulli distribution. More precisely, the channel realization $h(t)$ is equal to 1 if the packet is successfully decoded by the receiver side and is 0 otherwise. To that extent, we define the success probability as $\Pr(h(t)=1)=p_s$ and the failure probability as $\Pr(h(t)=0)=p_f=1-p_s$. We consider that when a packet is delivered to the receiver, the receiver sends an ACK packet back to the transmitter. In the case of a transmission failure, a Negative Acknowledgement (NACK) is sent by the receiver. We suppose that the ACK/NACK packets are instantaneously delivered to the transmitter [7, 84]. This assumption is widely used in the literature since the ACK/NACK packets are small. Accordingly, their transmission times can be considered negligible. Using these ACK/NACK packets, the transmitter can have perfect knowledge of the receiver's information source estimate at any time slot t.

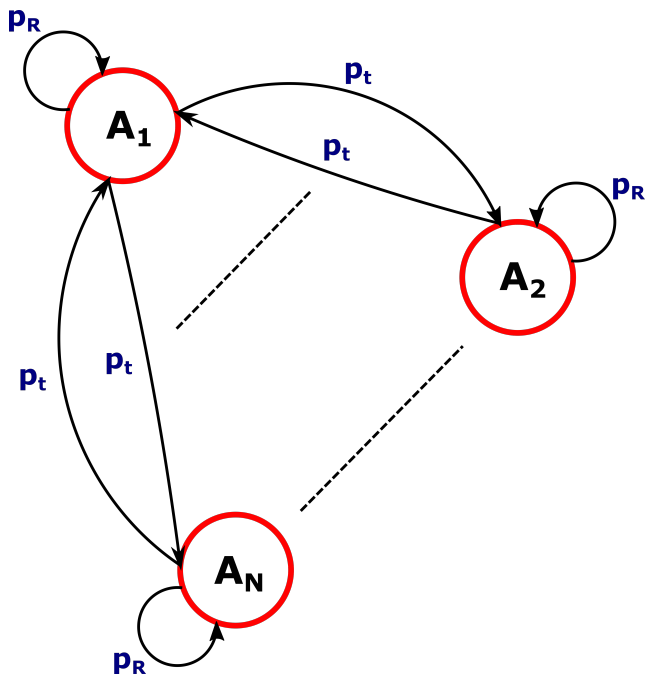


Figure 5.4: Illustration of the process of interest.

The next aspect of our model that we tackle is the nature of packets in the system. To that extent, we consider that the transmitter can generate information updates any time at its own will. More specifically, when the transmitter decides to send an update at time t, it samples $X(t)$ and proceeds to the transmission stage. If the packet is not successfully delivered to the receiver, and if the transmitter desires a transmission retrial at time $t+1$, a *new* status update is generated by sampling $X(t+1)$, and the transmission stage begins again.

Lastly, and as previously explained in the preceding section, the transmitter's ultimate objective is to adopt a transmission policy that minimizes the time average of a particular penalty function. In the sequel, we adopt the newly proposed metric reported in (5.10). To fully characterize it, we provide details on its dynamics in the next subsection.

5.3.2 Penalty Function Dynamics

For ease of notation, we will let $S(t)$ denote the special case of the AoII penalty previously reported in (5.10). More specifically:

$$S(t) = (t - V(t))\mathbb{1}\{\hat{X}(t) \neq X(t)\}, \tag{5.13}$$

where $V(t)$ is the last time instant where the monitor was in a correct state. In the sequel, we provide details concerning the dynamics of $S(t)$ to characterize the values of $S(t+1)$. To do so, we first define $\psi(t)$ as the decision at time t of the transmitter to either transmit (value 1) or to remain idle (value 0). We distinguish in the following between two cases: $S(t) = 0$ and $S(t) \neq 0$.

Case 1 - $S(t) = 0$: In this case, the monitor has perfect knowledge of the process of interest at time t. If the transmitter decides not to send a status update, then $S(t+1)$ will be equal to 0 if the process does not change value. This happens with a probability p_R. In the same fashion, $S(t+1)$ will be equal to 1 if the process changes value, which happens with a probability $1 - p_R = (N-1)p_t$. Let us now consider the case where the transmitter decides to send a status update at time t. Regardless of the channel realization, no new information will be conveyed to the monitor as $\hat{X}(t+1)$ will have the same value of $\hat{X}(t)$. Consequently, the previous analysis still holds for this case, and $S(t+1)$ will be equal to 0 if the process does not change value and 1 otherwise. We summarize what was stated in the following:

- $\Pr\big(S(t+1) = 0 | S(t) = 0, \psi(t) = 0\big) = \Pr\big(S(t+1) = 0 | S(t) = 0, \psi(t) = 1\big) = p_R$
- $\Pr\big(S(t+1) = 1 | S(t) = 0, \psi(t) = 0\big) = \Pr\big(S(t+1) = 1 | S(t) = 0, \psi(t) = 1\big) = 1 - p_R = (N-1)p_t$

Case 2 - $S(t) \neq 0$: In this case, the monitor does not have the correct knowledge of the process of interest (i.e., $\hat{X}(t) \neq X(t)$). If the transmitter decides to remain idle, then $S(t+1)$ will be equal to 0 if and only if the information process changes to the monitor's value from its last received update. More specifically, this is when $X(t+1) = X(U(t))$ with $U(t)$ being the time-stamp of the last successfully received packet by the monitor. This event occurs with a probability p_t. On the other hand, if the process keeps its same value, or transition to one of the remaining $N-2$ states, the penalty will grow by a step, i.e., $S(t+1) = S(t) + 1$. Now, let us consider the case where the transmitter decides to send a packet. To that extent, we consider two cases:

- $h(t) = 0$: In this case, the transmitted packet is not successfully decoded by the receiver. Therefore, no new knowledge is given to the monitor, i.e., $\hat{X}(t+1) = \hat{X}(t)$. Conditioned on $h(t) = 0$, we can assert that $S(t+1)$ becomes zero if and only if the information process changes to the value that the monitor has from its last received update. As previously mentioned, this event occurs with a probability p_t. On the other hand, $S(t+1)$ will be equal $S(t)+1$ if the process keeps its same value or change to one of the other $N-2$ states, which happens with a probability $p_R + (N-2)p_t$.
- $h(t) = 1$: In this case, the transmitted packet is successfully decoded by the receiver. Therefore, the estimate at the monitor $\hat{X}(t+1)$ is nothing but $X(t)$. To that extent, $S(t+1)$ will be equal to zero if the information process did not change during the transmission slot. This event happens with a probability p_R. On the other hand, if the process has changed during transmission to any of the remaining $N-1$ states, $S(t+1)$ will increase by 1.

By taking into account the independence between the information process transitions and the channel realizations, we can summarize the transitions probabilities of $S(t)$ in the following:

- $\Pr\big(S(t+1)=0|S(t)\neq 0,\psi(t)=0\big)=p_t$
- $\Pr\big(S(t+1)=S(t)+1|S(t)\neq 0,\psi(t)=0\big)=p_R+(N-2)p_t$
- $\Pr\big(S(t+1)=0|S(t)\neq 0,\psi(t)=1\big)=p_Rp_s+p_fp_t$
- $\Pr\big(S(t+1)=S(t)+1|S(t)\neq 0,\psi(t)=1\big)=p_Rp_f+(N-2)p_t+p_sp_t$

5.4 Unconstrained Scenario

5.4.1 Problem Formulation

Our aim is to find a transmission policy that minimizes the total average AoII of the network. A transmission policy ϕ is defined as a sequence of actions $\phi=(\psi^\phi(0),\psi^\phi(1),\ldots)$ where $\psi^\phi(t)=1$ if a transmission is initiated at time t. By letting Φ denote the set of all possible causal scheduling policies, our problem can be formulated as follows:

$$\underset{\phi\in\Phi}{\text{minimize}} \quad \lim_{T\to+\infty}\sup\frac{1}{T}\mathbb{E}^\phi\Big(\sum_{t=0}^{T-1}S^\phi(t)|S(0)\Big). \tag{5.14}$$

5.4.2 MDP Characterization

Based on our model's assumptions and the dynamics previously detailed in Section 5.3.2, our problem in (5.14) can be cast into an infinite horizon average cost Markov decision process that is defined as follows:

- **States**: The MDP state at time t is nothing but the penalty function $S(t)$. This penalty can have any value in $\mathbb{N}$. Therefore, the considered state space is countable and infinite.
- **Actions**: The action at time t, denoted by $\psi(t)$, indicates if a transmission is attempted (value 1) or the transmitter remains idle (value 0).
- **Transitions probabilities**: The transitions probabilities between the different states have been previously detailed in Section 5.3.2.
- **Cost**: We let the instantaneous cost of the MDP, $C(S(t),\psi(t))$, to be simply the penalty function $S(t)$.

Finding the optimal solution of an infinite horizon average cost MDP is recognized to be challenging due to the curse of dimensionality. More precisely, it is well-known that the optimal policy ϕ^* of the problem mentioned above can be obtained by solving the following Bellman equation [85]:

$$\theta+J(S)=\min_{\psi\in\{0,1\}}\Big\{S+\sum_{S'\in\mathbb{N}}\Pr(S\to S'|\psi)J(S')\Big\},\quad \forall S\in\mathbb{N}, \tag{5.15}$$

where $\Pr(S \rightarrow S'|\psi)$ is the transition probability from state S to S' given the action ψ, θ is the optimal value of (5.14) and $J(S)$ is the differential cost-to-go function. Based on (5.15), one can see that the optimal policy ϕ^* depends on $J(\cdot)$, for which there is no closed-form solution in general [85]. There exist various numerical algorithms in the literature that solve (5.15), such as the value iteration and the policy iteration algorithms. However, they suffer from being computationally demanding. To circumvent this complexity, we study in the next section the structural properties of the optimal transmission policy.

5.4.3 Structural Results

The first step in our structural analysis of the optimal policy consists of studying the particularity of the function $J(.)$. To that extent, we provide the following lemma.

Lemma 5.1. *The value function $J(S)$ is increasing in S.*

Proof. The proof can be found in Appendix C.1. □

The above lemma will be used in the following theorem to provide results on the optimal transmission policy.

Theorem 5.1. *The optimal transmission policy ϕ^* of our problem in (5.14) is:*

- *$p_t < p_R$: the transmitter should send updates at each time slot or when the receiver is in an erroneous state. In both cases, the optimal cost is:*

$$\overline{C}_{AU} = (N-1)p_t \frac{\frac{1}{(1-a)^2}}{1 + \frac{(N-1)p_t}{1-a}}. \tag{5.16}$$

- *$p_t \geq p_R$: it is optimal to never transmit any packet. In this case, the optimal cost is:*

$$\overline{C}_{NU} = \frac{(N-1)p_t}{(1-b)^2 + (1-b)(N-1)p_t}. \tag{5.17}$$

with a, b being two constants that are equal to $p_R p_f + (N-2)p_t + p_s p_t$ and $p_R + (N-2)p_t$ respectively.

Proof. The proof can be found in Appendix C.2. □

The intuition behind the above results is that when $p_t \geq p_R$, a transmitted packet has a high chance of becoming erroneous by the time it is delivered to the receiver. Accordingly, in this case, the information source changes so fast to the point that transmitting packets will harm the performance of the system. As this case is not of practical interest, we focus in the rest of the chapter on the scenario where $p_t < p_R$. Consequently, in the case where no constraints on the power are imposed, the minimum cost is achieved either by sending updates at every time slot or when the receiver is in an erroneous state.

Remark 5.1. *By adopting the same model as the one above, and by considering the AoI as the penalty function, one can show that the optimal transmission policy is to send updates at each time slot. As for the error penalty function, one can verify that sending updates at every time slot or when the receiver is in an erroneous case minimizes the prediction error. Consequently, and as the intuition suggests, an "always update" policy minimizes all the above* 3 *penalties in the unconstrained power case. However, as shown in the sequel, this does not hold in the case of power-constrained scenarios.*

5.5 Power-Constrained Scenario

5.5.1 Problem Formulation

In realistic scenarios, a transmitter cannot send status updates at each time slot. Each attempted transmission incurs a power cost δ, and the transmitter has an average power budget δ_{budget} that cannot be surpassed. Consequently, the transmitter must wisely choose when to transmit an update to the monitor since the following constraint has to be satisfied by any chosen transmission policy ϕ:

$$\lim_{T\to+\infty} \sup \frac{1}{T}\mathbb{E}^{\phi}\Big(\sum_{t=0}^{T-1} \delta\psi^{\phi}(t)\Big) \leq \delta_{\text{budget}}, \tag{5.18}$$

where the transmission policy ϕ is defined as a sequence of actions $\phi = (\psi^{\phi}(0), \psi^{\phi}(1), \ldots)$ such that $\psi^{\phi}(t) = 1$ if a transmission is initiated at time t. Since $\lim_{T\to+\infty} \sup \frac{1}{T}\mathbb{E}^{\phi}\Big(\sum_{t=0}^{T-1} \psi^{\phi}(t)\Big) \leq 1$, we define $\alpha = \frac{\delta_{\text{budget}}}{\delta}$ and we suppose that $\alpha \leq 1$ as the constraint becomes redundant otherwise. Putting it all together, our problem can be formulated as follows:

$$\begin{aligned} \underset{\phi\in\Phi}{\text{minimize}} \quad & \lim_{T\to+\infty} \sup \frac{1}{T}\mathbb{E}^{\phi}\Big(\sum_{t=0}^{T-1} S^{\phi}(t)|S(0)\Big), \\ \text{subject to} \quad & \lim_{T\to+\infty} \sup \frac{1}{T}\mathbb{E}^{\phi}\Big(\sum_{t=0}^{T-1} \psi^{\phi}(t)\Big) \leq \alpha. \end{aligned} \tag{5.19}$$

To address the above problem, we proceed with a Lagrange approach that transforms our constrained minimization problem into an optimization of the Lagrangian function. More specifically, by letting $\lambda \in \mathbb{R}^{+}$ be the Lagrange multiplier, we define the Lagrangian function as follows:

$$f(\lambda, \phi) = \lim_{T\to+\infty} \sup \frac{1}{T}\mathbb{E}^{\phi}\Big(\sum_{t=0}^{T-1} S^{\phi}(t) + \lambda\psi^{\phi}(t)|S(0)\Big) - \lambda\alpha. \tag{5.20}$$

To that end, the Lagrange approach can be summarized in the following problem:

$$\max_{\lambda\in\mathbb{R}^{+}} \min_{\phi\in\Phi} f(\lambda, \phi). \tag{5.21}$$

It is well-known that for any feasible scheduling policy ϕ, the optimal value of the problem in (5.21) forms a lower bound to that of our original problem in (5.19) [86]. The difference between the two values is known as the *duality* gap, which is generally non-zero. Our goal is to show that our approach can achieve the optimal solution of the problem in (5.19). To that extent, we first study in the sequel the problem:

$$g(\lambda) = \min_{\phi \in \Phi} \; f(\lambda, \phi). \tag{5.22}$$

5.5.2 MDP Characterization

Similarly to the previous section, we cast the problem (5.22) into an MDP, which is the same as the one reported in the previous section except for the cost being defined in this case as:

$$C(S(t), \psi(t)) = S(t) + \lambda \psi(t). \tag{5.23}$$

Following the same line of work, we know that the optimal policy ϕ^* of the problem $\min_{\phi \in \Phi} \; f(\lambda, \phi)$ can be obtained by solving the Bellman equation for all $S \in \mathbb{N}$:

$$\theta' + J'(S) = \min_{\psi \in \{0,1\}} \Big\{ S + \lambda \psi + \sum_{S' \in \mathbb{N}} \Pr(S \to S' | \psi) J'(S') \Big\}, \tag{5.24}$$

where $\Pr(S \to S'|\psi)$ is the transition probability from state S to S' given the action ψ, θ' is the optimal value of the problem and $J'(S)$ is the differential cost-to-go function. As detailed in the previous section, solving the above equation directly is cumbersome in complexity. Hence, we provide structural properties of the optimal transmission policy in the next subsection.

5.5.3 Structural Results

In the same spirit as the previous section, we start by investigating the properties of the function $J'(.)$.

Lemma 5.2. *The function $J'(S)$ is increasing in S.*

Proof. The proof follows the same procedure of Lemma 5.1 and is therefore omitted. □

The above lemma will be used to show that the optimal policy of our problem is a threshold policy. Before providing the proof of our claim, we first lay out the following definition.

Definition 5.1. *An increasing threshold policy is a deterministic stationary policy in which the transmitter remains idle if the current state of the system S is smaller than n and attempts to transmit otherwise. In this case, the policy is fully characterized by the threshold $n \in \mathbb{N}$.*

With the above definition being laid out, we present the following proposition.

Proposition 5.1. *The optimal policy ϕ^* of the problem in (5.22) is an increasing threshold policy.*

Proof. The proof can be found in Appendix C.3. □

With the structure of the optimal policy of (5.22) being found, we tackle in more depth the average cost of our MDP when a threshold policy is adopted. To that extent, we recall that a threshold policy is fully characterized by its threshold value n. Accordingly, our problem in (5.22) can be reformulated as follows:

$$\underset{n\in\mathbb{N}}{\text{minimize}} \quad \overline{C}(n,\lambda), \tag{5.25}$$

where $\overline{C}(n,\lambda)$ is the infinite horizon average cost of the MDP when the threshold policy is adopted. To find the expression of $\overline{C}(n,\lambda)$, $\forall n \in \mathbb{N}$, we first tackle the special case where the transmitter always sends updates at each time slot (i.e., $n = 0$). In this scenario, the portion of time where the transmitter is sending updates, which is defined as $\lim_{T\to+\infty} \sup \frac{1}{T}\mathbb{E}^{\phi}\Big(\sum_{t=0}^{T-1} \psi^{\phi}(t)\Big)$, is equal to 1. Moreover, by using Theorem 5.1, we end up with the following:

$$\overline{C}(0,\lambda) = (N-1)p_t \frac{\frac{1}{(1-a)^2}}{1+\frac{(N-1)p_t}{1-a}} + \lambda(1-\alpha), \tag{5.26}$$

with a being equal to $p_R p_f + (N-2)p_t + p_s p_t$. Next, we shift our attention to the case where $n \in \mathbb{N}^*$. To that extent, we note that for any threshold policy, the MDP can be modeled through a Discrete Time Markov Chain (DTMC) where:

- The states refer to the values of the AoII $S(t)$.
- For any state $S(t) < n$, the transmitter is idle and therefore the dynamics of $S(t)$ coincide with those of $\psi(t) = 0$ of Section 5.3.2. On the other hand, for any state $S(t) \geq n$, the dynamics of $S(t)$ coincide with those of $\psi(t) = 1$ of the same section.

Consequently, we focus in the sequel on this DTMC.

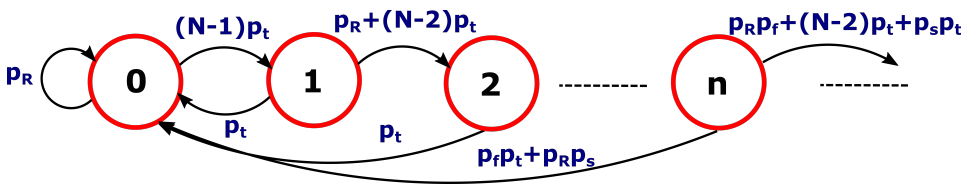


Figure 5.5: The states transitions under a threshold policy.

The next step towards finding the average cost $\overline{C}(n,\lambda)$ consists of calculating the DTMC's stationary distribution. We, therefore, provide the following proposition.

Proposition 5.2. *For a fixed threshold $n \in \mathbb{N}^*$, the DTMC in question is irreducible and admits $\pi_k(n),\ \forall k \in \mathbb{N}$ as its stationary distribution where:*

$$\pi_0(n) = \frac{1}{1 + \frac{(N-1)p_t(1-b^n)}{1-b} + \frac{(N-1)p_t a b^{n-1}}{1-a}}, \tag{5.27}$$

$$\pi_k(n) = (N-1)p_t b^{k-1}\pi_0, \quad 1 \le k \le n, \tag{5.28}$$

$$\pi_k(n) = (N-1)p_t b^{n-1} a^{k-n}\pi_0, \quad k \ge n+1, \tag{5.29}$$

with a, b being two constants that are equal to $p_R p_f + (N-2)p_t + p_s p_t$ and $p_R + (N-2)p_t$ respectively.

Proof. The proof can be found in Appendix C.4. □

By leveraging Proposition 5.2, we can proceed to find a closed-form of the average cost of the threshold policy.

Theorem 5.2. *For a fixed threshold $n \in \mathbb{N}^*$, the average cost of the policy is $\overline{C}(n,\lambda) = \overline{C}(n) + \overline{C}_1(n,\lambda)$ where:*

$$\overline{C}(n) = (N-1)p_t \frac{\frac{1+b^n(nb-n-1)}{(1-b)^2} + \frac{b^{n-1}a(n+\frac{1}{1-a})}{1-a}}{1 + \frac{(N-1)p_t(1-b^n)}{1-b} + \frac{(N-1)p_t a b^{n-1}}{1-a}}, \tag{5.30}$$

$$\overline{C}_1(n,\lambda) = \lambda \frac{(N-1)p_t b^{n-1}}{(1-a)(1 + \frac{(N-1)p_t(1-b^n)}{1-b} + \frac{(N-1)p_t a b^{n-1}}{1-a})} - \lambda\alpha. \tag{5.31}$$

Proof. The proof can be found in Appendix C.5. □

As we now have the expression of the average cost $\overline{C}(n,\lambda)$, we turn our attention to studying its characteristics in order to prove the optimality of the Lagrange approach.

5.5.4 Optimality of the Lagrange Approach

The Lagrange approach's optimality in similar resource-constrained environments has been established in the literature for other cost functions (e.g., the AoI in [23]). However, contrary to [23], the standard approach to prove this optimality cannot be adopted in our case. This is mainly due to the complexity of the average cost function reported in Theorem 5.2. In particular, as seen in (5.30)-(5.31), $\overline{C}(n,\lambda)$ is not necessarily convex in n, which limits the applicability of the approach adopted in [23]. Accordingly, to demonstrate the Lagrange approach's optimality in our case, we proceed in a different direction. Specifically, we investigate the behavior

of the cost function in more depth and leverage these results to establish the AoII-optimal policy. To present our approach, we fix the threshold n and we let

$$A(n) = \sum_{k=n}^{+\infty} \pi_k(n), \quad \forall n \in \mathbb{N} \tag{5.32}$$

denote the portion of time where the transmitter is attempting to send packets. By definition, we have that the sequence $\big(A(n)\big)_{n\in\mathbb{N}}$ is decreasing with $A(0) = 1$. Using the expressions of $\pi_k(n)$ in Proposition 5.2, we can obtain

$$A(n) = \frac{(N-1)p_t b^{n-1}}{(1-a)(1+\frac{(N-1)p_t(1-b^n)}{1-b}+\frac{(N-1)p_t a b^{n-1}}{1-a})}, \quad \forall n \in \mathbb{N}^*. \tag{5.33}$$

To that end, we have $\overline{C}_1(n,\lambda) = \lambda A(n) - \lambda\alpha$. With this definition in mind, we summarize our approach in the following:

1. We prove that $\overline{C}(n)$, which is reported in (5.30), is increasing with n.
2. We define the set of intersection points $\lambda(n)$ as the values of λ such that $\overline{C}(n,\lambda) = \overline{C}(n+1,\lambda)$. Accordingly, we have
$$\lambda(n) = \frac{\overline{C}(n+1) - \overline{C}(n)}{A(n) - A(n+1)}, \quad \forall n \in \mathbb{N}. \tag{5.34}$$
3. We prove that $\lambda(n)$ is increasing with n.
4. We relate through graphical methods and several inductive lemmas the results on $\lambda(n)$ to the establishment of the AoII-optimal policy.
5. We propose a low-complexity algorithm to find the AoII-optimal operating point of the system.

The details of the above steps will be laid out in the remainder of this section. To proceed in this direction, we also define $n(\lambda)$ as the optimum threshold that solves, for a fixed λ, the optimization problem (5.25). With these definitions, and with our steps being clarified, we now proceed with the proof of optimality. To that extent, let us first note that the following always holds:

$$g(\lambda) \leq \max_{\lambda\in\mathbb{R}^+} g(\lambda) \leq \theta^*, \tag{5.35}$$

where θ^* is the optimal value of our constrained problem (5.19). To understand this inequality, we recall that $\max_{\lambda\in\mathbb{R}^+} g(\lambda)$ forms a lower bound for θ^*. Therefore, given that $g(\lambda) \leq \max_{\lambda\in\mathbb{R}^+} g(\lambda)$, we can deduce the inequality in (5.35). Consequently, it is evident that if we can find λ_1 such that $A(n(\lambda_1)) = \alpha$, then $g(\lambda_1) = \max_{\lambda\in\mathbb{R}^+} g(\lambda) = \theta^*$. In this case, we achieve the optimal operating point of (5.19) by simply adopting a threshold policy characterized by the threshold $n(\lambda_1)$. However, the issue arises when such a value of λ_1 does not exist since the set $\{n(\lambda) : \lambda \in \mathbb{R}^+\}$ is discrete. To deal with this, we aim to show that, thanks to the properties of the AoII function, we can always find (n_0, λ_{n_0}) such that:

1. $\overline{C}(n_0, \lambda_{n_0}) = \overline{C}(n_0 + 1, \lambda_{n_0})$,
2. $\begin{cases} A(n_0) \geq \alpha, \\ A(n_0 + 1) < \alpha, \end{cases}$
3. $n(\lambda_{n_0}) = n_0$.

In this case, it is sufficient to take a mixture of two threshold policies ϕ_{n_0} and ϕ_{n_0+1} with a probability $\rho = \frac{\alpha - A(n_0+1)}{A(n_0) - A(n_0+1)}$ and $1-\rho = \frac{A(n_0) - \alpha}{A(n_0) - A(n_0+1)}$ respectively, to achieve the optimal objective value of the constrained problem in (5.19). We now proceed to show the existence and uniqueness of (n_0, λ_{n_0}).

Proposition 5.3. *The following always holds:*

$$\forall n \in \mathbb{N}, \exists \lambda_n \in \mathbb{R}^+ : \overline{C}(n, \lambda_n) = \overline{C}(n+1, \lambda_n). \tag{5.36}$$

Proof. The proof can be found in Appendix C.6. □

As the above proposition holds for any n, let us focus on the value n_0 such that:

$$\begin{cases} A(n_0) \geq \alpha, \\ A(n_0 + 1) < \alpha. \end{cases} \tag{5.37}$$

In the next theorem, we show that this value n_0 verifies $n(\lambda_{n_0}) = n_0$.

Theorem 5.3. *For the aforementioned λ_{n_0}, n_0 minimizes the average cost function $\overline{C}(n, \lambda_{n_0})$.*

Proof. The proof can be found in Appendix C.7. □

5.5.5 Algorithm Implementation

Based on the previous section, we can assert that the optimal transmission policy is a mixture of two deterministic threshold policies ϕ_{n_0} and ϕ_{n_0+1} such that:

$$\begin{cases} A(n_0) \geq \alpha, \\ A(n_0 + 1) < \alpha. \end{cases} \tag{5.38}$$

As $\big(A(n)\big)_{n \in \mathbb{N}}$ is a decreasing sequence in n, we can rewrite $n' = n_0 + 1$ as follows:

$$n' = \inf\{n \geq 1 : A(n) - \alpha < 0\}. \tag{5.39}$$

For any $0 < \alpha \leq 1$, we can attest that there exists a finite n' that verifies the above condition. To find this value, we employ a two steps algorithm depicted in Algorithm 2. The two steps are as follows:

Algorithm 2 Optimal threshold finder

procedure UPPERBOUND INCREASE
  Init. $N_{LB} = N_{UB} = 1$
  while $A(N_{UB}) - \alpha \geq 0$ **do**
    $N_{LB} := N_{UB}$
    $N_{UB} := 2N_{UB}$
  end while
end procedure
procedure BINARY SEARCH
  $n' := \left\lceil \frac{N_{LB}+N_{UB}}{2} \right\rceil$
  while $n' < N_{UB}$ **do**
    if $A(n') - \alpha \geq 0$ **then** $N_{LB} := n'$
    else $N_{UB} = n'$
    end if
    $n' := \left\lceil \frac{N_{LB}+N_{UB}}{2} \right\rceil$
  end while
end procedure
Output the optimal threshold $n_0 = n' - 1$

- Exponential increase of the upperbound value N_{UB} to ensure that n' is included in the interval of interest $[N_{LB}, N_{UB}]$.
- A binary search in the interval mentioned above to find the value n'.

The first part of the algorithm finishes in $N_1 = \log_2(n')$ iterations while the binary search part is known to have a worst-case complexity of $\log_2(N_{size})$ where N_{size} is the size of the interval of interest. To that extent, we have that: $N_{size} = 2^{N_1} - 2^{N_1-1}$. Hence, the worst-case complexity of the second part is $\log_2(N_{size}) = N_1 - 1$. Therefore, we can conclude that the complexity of the above algorithm is logarithmic in the value of n', which makes it appealing to be implemented in practice.

After the algorithm finishes and n_0 is found, we adopt a transmission policy where a packet is generated and transmitted when the penalty is equal to n_0 and $n_0 + 1$ with a probability $\rho = \frac{\alpha - A(n_0+1)}{A(n_0) - A(n_0+1)}$ and $1 - \rho = \frac{A(n_0) - \alpha}{A(n_0) - A(n_0+1)}$ respectively. By doing so, we achieve the optimal objective value of the constrained problem in (5.19).

5.6 Numerical Results

In this section, we provide numerical results that highlight the effects of the information source dynamics on the performance of our proposed AoII-optimal policy. We also compare our framework to both the AoI and the error function minimization frameworks in order to shed light on important insights. Note that, although

we focus on the Markovian information source depicted in Section 5.3.1, the insights provided in this section intuitively hold for more general information source models.

5.6.1 Information Source Parameters

In the first scenario, we investigate in more depth the effect of the Markov chain's dynamics on the performance of our proposed AoII-optimal policy.

Effect of p_R

In this scenario, we consider that the number of states is $N = 8$, and we fix the parameter α to 0.1. As for the channel parameter, we assume that the transmission success probability p_s is equal to 0.8. While making sure that $p_t < p_R$, we vary the probability of remaining in the same state p_R and plot the average AoII of the optimal policy. As seen in Fig. 5.6, the average cost decreases as p_R increases. The reason behind this is twofold:

1. When p_R is high, the information source becomes more "predictable". In other words, when a packet is transmitted, it is less likely for it to become obsolete due to a transition of the Markov chain during the transmission stage.

2. When p_R is high, the AoII remains zero for a significant amount of time upon successful transmission. This allows us to make better use of the permitted power budget α as we will be able to transmit at a lower threshold value without exceeding the allowed power budget. This can be verified by looking at n_0 in function of p_R in the following table:

p_R	0.2	0.4	0.6	0.8
n_0	15	12	10	7

Table 5.1: Variation of n_0 in function of p_R.

From the above table, we can see that as p_R increases, the value of n_0 decreases. In other words, our tolerance for the AoII is reduced, and we can transmit at a much lower AoII value without violating the power constraint. This eventually leads to a reduction in the average AoII.

Effect of N

We consider the case where $p_R = 0.5$, $\alpha = 0.1$, and the probability of successful transmission is $p_s = 0.8$. We vary N and report the average AoII when the AoII-optimal policy is adopted in Fig. 5.7. As shown in the figure, the average AoII

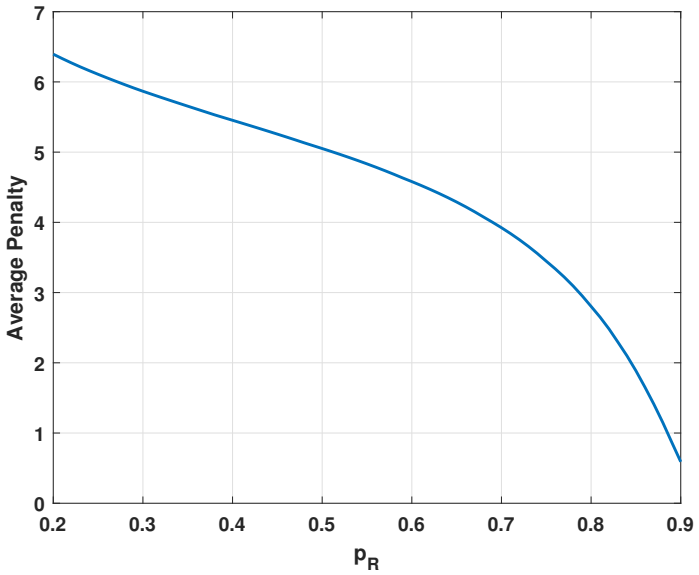


Figure 5.6: The average AoII in function of p_R.

increases when the number of states N grows. To explain this trend, we first recall that the transition probabilities at each state always verify the following equality:

$$p_R + (N-1)p_t = 1 \tag{5.40}$$

Accordingly, we can use (5.40) to conclude that $p_t = \frac{1-p_R}{N-1}$. Next, let us consider that the monitor has perfect knowledge of the information process at time t, denoted by $X(t)$. Then, let us suppose that the information source changes value at time $t+1$, which happens with a fixed probability $1-p_R$. With that in mind, we recall that the probability for the information source to go back to its old value $X(t)$ at time $t+2$ is p_t. As p_t is a decreasing function with N, this means that the probability for the monitor to have correct knowledge of the information source at time $t+2$ without wasting resources for packet transmission decreases with N. Accordingly, when N grows, the average AoII will also increase.

5.6.2 Comparison with the AoI Framework

In the following, we provide a comparison between our optimal transmission policy and the optimal age policy of [23].

Comparison in Function of α

In this case, we adopt the same number of states $N = 8$ and success probability $p_s = 0.8$. We fix the probability of remaining in the same state p_R to 0.5. We vary the parameter α and plot the average AoII achieved by both policies. As seen in Fig. 5.8, the proposed policy always outperforms the age-optimal policy for all values of α. The following two observations can also be drawn from the figure:

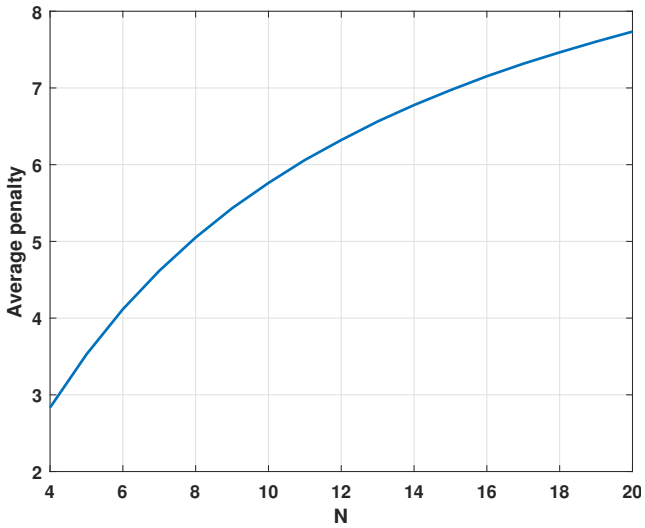


Figure 5.7: The average AoII in function of N.

1. One can see that the two curves converge as α increases. This is in agreement with our theoretical results in the unconstrained case in Section 5.4. In fact, when the imposed power constraint becomes less restrictive, the transmitter will be sending more packets and we converge to the "always update" policy, which minimizes both the AoII and the AoI.

2. Another interesting observation is that the gap between the two curves is small when α is small (e.g., the gap is equal to 1.1 for $\alpha = 0.02$). This is due to the number of packets sent by the transmitter becoming very small. Consequently, the average AoII will be mostly dictated by how the Markov chain evolves rather than the transmission policy adopted. Therefore, in this case, we converge to the "no updates" average cost previously reported in eq. (5.17).

By combining the above two observations, we can conclude that when the transmitter is 1) severely constrained by its power or 2) has unlimited power, age-optimal policies lead to virtually the same performance as the optimal AoII policy.

We also investigate the age performance of our proposed policy and compare it to the age-optimal policy. As seen in Fig. 5.9, the age-optimal policy outperforms our policy in terms of average age. However, the gap between the two curves vanishes for high α, and that is for the same reason previously reported in the average AoII comparison between the two policies. On the other hand, as α decreases, the gap between the two curves increases, reaching 190 for $\alpha = 0.02$. The reason behind this is the fact that as α decreases, the allowed number of transmissions becomes extremely small. Therefore, the impact of the transmission decisions will become more significant on the performance. To that extent, since our policy is based on the information content of the packet rather than just the age at the monitor, our proposed penalty measure can sometimes be equal to 0 while the age is

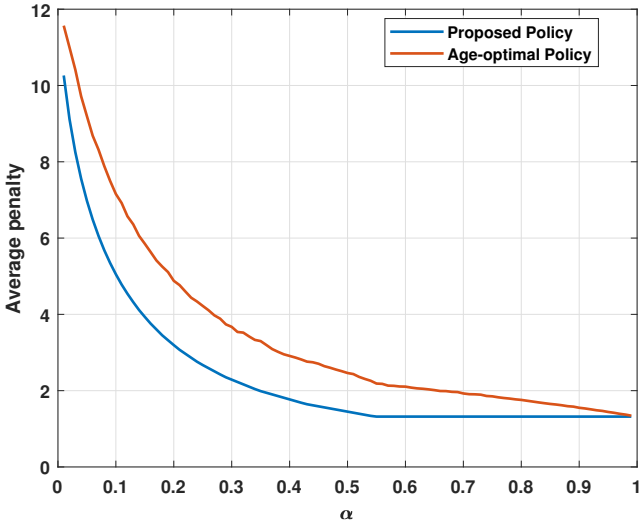


Figure 5.8: Comparison between our proposed policy and the age-optimal transmission policy in terms of average AoII.

equal to 100. These differences between the two transmission policies will lead to a significant gap in age performance when the available power budget is tiny.

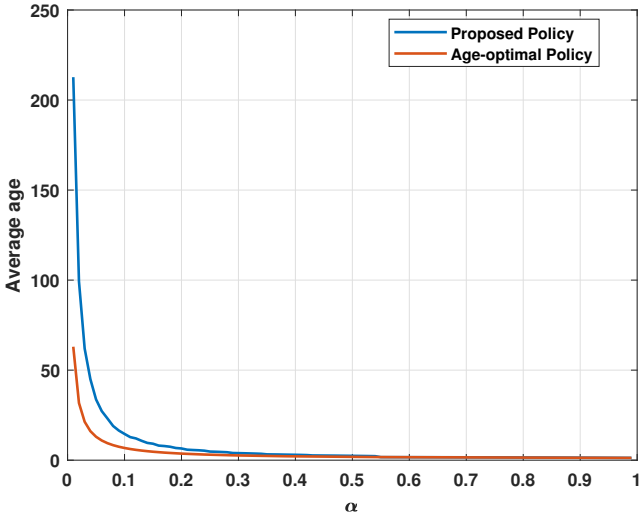


Figure 5.9: Comparison between the two policies in terms of average age.

Comparison in Function of p_R

In this scenario, we compare the AoII-optimal policy and the AoI-optimal policy when p_R is varied. We consider that $N = 8$, $\alpha = 0.1$, and the probability of successful transmission is $p_s = 0.8$. While maintaining $p_t < p_R$, we vary p_R and report the differences between the two policies in Fig. 5.10. As can be seen in the figure, the gap between the two curves increases as p_R grows (from 0.7 for $p_R = 0.2$ to 2.2 for $p_R = 0.9$). To explain this, we recall that the AoI always increases regardless of the information source's value. As p_R increases, the information process $X(t)$ will have

a higher probability of keeping the same value at the next time slot $t + 1$. However, since the AoI is always increasing, the AoI-optimal policy will waste vital resources to update the monitor when it is not necessary to do so. As the AoI-optimal policy sends more obsolete packets when p_R is high, this will to a non-negligible gap between the AoI-optimal and AoII-optimal policies as seen in the figure.

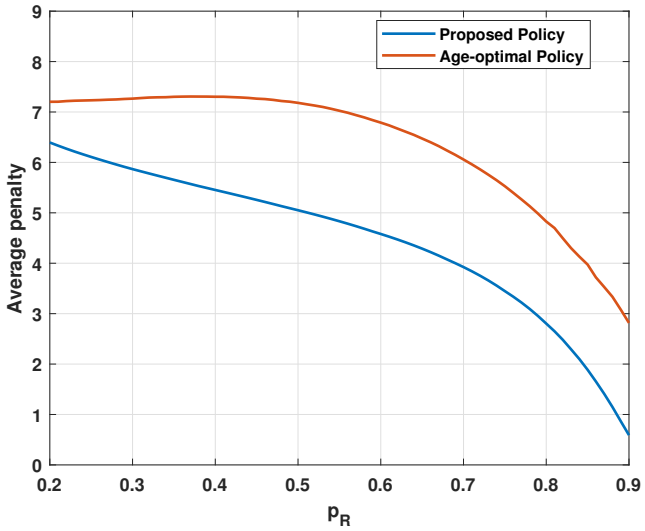


Figure 5.10: Comparison in function of p_R between the AoII-optimal policy and the AoI-optimal policy.

5.6.3 Comparison with the Error Framework

In the following, we present a comparison between our policy and the error-based policy that follows the rules below:

- Send a packet solely when the monitor has a wrong estimate of the information source.
- Ensure that the constraint on the power consumption is verified with *equality*.

We consider the case where $N = 8$, $p_R = 0.5$, and the probability of successful transmission is $p_s = 0.8$. We report in the next table the AoII values of the two policies. As can be seen in Table 5.2, our policy always outperforms the error-based policy. We can also see that as α increases, the gap between the two shrinks since the transmitter will be sending more packets, and we converge to the “always update" policy, which minimizes both the AoII and the status error function.

α	$\text{AoII}_{\text{optimal}}$	$\text{AoII}_{\text{error}}$
0.12	4.4	5.9
0.25	2.7	3.8
0.45	2	2.2

Table 5.2: Comparison between the AoII-optimal policy and the error based policy.

5.6.4 Real-Life Application of the AoII Framework

To put into perspective our proposed framework, we consider a transmitter-receiver pair where a real-time video stream packets are sent from one end of the network to the other. Time is considered to be slotted and normalized to the slot duration (i.e., the slot duration is taken as 1). The video stream comprises frames, each of which is a 1-D vector of length M in line-scan order. At each time slot, a frame of the video stream is sent by the transmitter side. We suppose that the network channel at time slot t is $X(t)$, and its estimate at the transmitter's side at the same time slot is $\hat{X}(t)$. For simplicity, we consider that the channel can only have two possible states and that the transitions between them happen with probability $0 < p_t < 0.5$. We assume that the receiver successfully decodes packets if $X(t) = \hat{X}(t)$, and a transmission error occurs otherwise. The transmitter can send pilot signals and learn the channel at the beginning of each time slot. However, this training succeeds with a probability $0 < p_s < 1$ and incurs a cost δ, knowing that an average cost budget δ_{budget} cannot be surpassed. This channel model is similar to the standard Gilbert–Elliott channel model, widely used in the wireless communications literature [87].

The transmitter's goal is to minimize the total distortion of the video signal at the receiver by learning the channel state. At the receiver, we assume a simple loss concealment scheme where the lost frame due to a transmission error is replaced by the previous frame at the decoder output. The error propagation process is modeled with a geometric attenuation factor resulting from spatial filtering. To simplify the model, we suppose that both the erroneous frames and the effects these losses have on the distortion are independent. To that end, and as derived in [88], the distortion of the video signal due to a transmission error burst of duration $B \geq 2$ time slots starting at frame index $k - B + 1$ is:

$$D(B) = \sum_{i=k-B+1}^{k-1} d[i] + \alpha(B) d[k], \tag{5.41}$$

where $d[i]$ is the MSE corresponding to the error frame i and $\alpha(B)$ is the ratio of distortion. It was found in [88] that the ratio of distortion is a near-linear function of B. In particular:

$$\alpha(B) = \alpha_0 + c(B - 2), \tag{5.42}$$

where α_0 is the ratio for $B = 2$, c is the slope of the increase, and $B \geq 2$. Accordingly, if we assume that the MSE corresponding to the error frame is the same for all lost frames, we end up with:

$$D(B) = (B - 2)d(1 + c) + \alpha_0 d. \tag{5.43}$$

Note that the distortion caused by an error of length $B = 1$ is d. To that end, the total average long-term distortion of the video signal at the receiver is:

$$D = \frac{1}{T} \sum_{i \in \Xi} D(B_i), \tag{5.44}$$

where Ξ is the set of indices of transmission errors, B_i is the duration of the i-th error, and T is the video stream duration. One can clearly see that the expression in (5.44) is a special case of the AoII where $f(t) = (t - V(t) - 2)d(1 + c) + \alpha_0 d$ and $g(X(t), \hat{X}(t)) = \mathbb{1}\{\hat{X}(t) \neq X(t)\}$. Consequently, one can see that this real-life application falls under the general AoII minimization framework reported in this chapter.

To highlight our AoII approach's benefits, we compare it to the AoI and the standard error-based frameworks for this particular scenario. Specifically, we evaluate the average video distortion that results from adopting the optimal policies for these 3 different metrics reported in (5.1)-(5.3). We consider a scenario where $\alpha_0 = 6$, $c = 2$, $d = 17$, $p_t = 0.4$, $p_s = 0.8$, and the length of the video $T = 10^6$ time slots. We vary the power budget ratio $\frac{\delta_{\text{budget}}}{\delta}$ and reports our findings in Fig. 5.11.

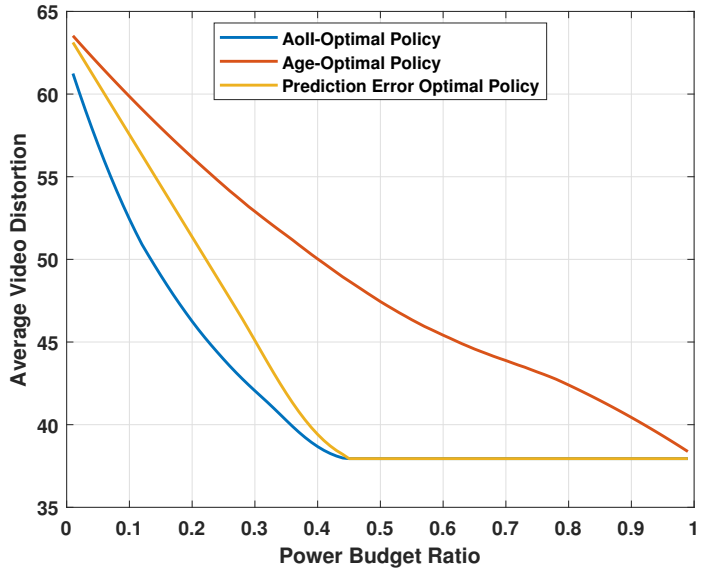


Figure 5.11: Comparison of the 3 optimal policies in function of the power budget ratio $\frac{\delta_{\text{budget}}}{\delta}$.

As seen in Fig. 5.11, the AoII-optimal policy outperforms the two other policies for the whole power budget ratio range. The performance gap between the policies shrinks as $\frac{\delta_{\text{budget}}}{\delta}$ increases. This is because as the transmitter becomes less restricted by the training cost, these policies converge to the "always update" policy where channel estimation happens at each time slot. This highlights the advantages of adopting the AoII framework and opens new perspectives for further real-life applications where the AoII can be of interest.

5.7 Conclusion

In this chapter, we have proposed a new performance metric that deals with the shortcomings of the conventional AoI and error penalty functions in the framework of status updates. Dubbed as the age of incorrect information, this new metric extends the notion of fresh updates and adequately captures the information content that the updates bring to the monitor. We have studied the metric mentioned above in the case where a transmitter-receiver pair communicates over an unreliable channel. By leveraging MDP tools, the optimal policy's structure was found for the cases where the transmitter is limited and non-limited by its power. A low-complexity algorithm was then presented that finds the optimal operating point that minimizes the average AoII. Lastly, numerical results were laid out that highlight the effect of the information source's dynamics on the AoII, along with a comparison between the AoI and AoII frameworks.

Part III

Conclusions, Outlook and Appendices

6 | Conclusions and Outlook

6.1 Conclusion

In this thesis, we have focused on investigating the notion of Age of Information in various network settings and system environments. In particular, we have studied, in Chapter 3, a transmission scheduling problem in which several streams of status update packets with diverse priority levels are sent through a common channel. After deriving each stream's average age for Poisson arrivals and exponential service times, we provided insights that suggest the need for a new age-optimization framework in multi-class systems. Accordingly, we have introduced the notion of lex-age-optimality to evaluate the performance of multi-class status update policies. To that end, we proposed a scheduling policy that we have shown to be lex-age-optimal for (i) minimizing any time-dependent, symmetric, and non-decreasing age penalty function; (ii) minimizing any non-decreasing functional of the stochastic process formed by the age penalty function; and (iii) the cases where different priority classes have distinct arrival traffic patterns, age penalty functions, and age penalty functionals.

In Chapter 4, we have examined a random access network where N links contend for the channel using the well-known CSMA scheme. Using SHS theory, we have found each link's average age in several packet arrival scenarios. Accordingly, we have studied the average age minimization problem and found the optimal back-off durations that solve this problem. We have also proposed a modified version of the standard CSMA scheme to ameliorate the network's age performance.

In the second part of the thesis, specifically in Chapter 5, we have shed light on the shortcomings of the age of information and standard error metrics in many real-time applications. To that end, we have introduced the age of incorrect information metric in the status updates framework. After motivating the metric, we have derived optimal sampling policies that minimize the AoII in both the unconstrained case and the case where an update rate cannot be exceeded. The proposed measure was then shown to have significant advantages over the AoI and the error penalty functions in various practical scenarios.

6.2 Future Research Directions

Although the analysis presented in this thesis enables a considerable understanding of the freshness of communication in various system settings, there exist several research directions that can be examined in the future.

Multi-Class Systems

In Chapter 3, we have investigated a multi-class system where the updates are sent through a common service facility. The facility was modeled a single server with exponential service times. The model and analysis of this chapter can be extended by considering

- *General service time distributions*: Although the exponential distribution is widely used in the literature, there exist various other prevalent distributions. Among the possible distributions that can be examined in the future are the constant time distribution, geometric distribution, and Erlang distribution. Compared to our continuous-time analysis, one may need to recourse to different mathematical tools for analyzing the discrete systems. For example, the SHS approach we adopted can be challenging to follow in such systems.

- *Multi-server scenarios*: The multi-server case is of ample importance as, ordinarily, the service facility can have more than one resource at its disposal. An interesting question that arises in this type of scenario is whether or not to replicate packets across the numerous servers to achieve the lex-age-optimality [89].

- *Multi-hop networks*: Age-optimal scheduling policies for single class multi-hop scenarios have been previously found in [17]. An appealing direction is to investigate the multi-class environment to find lex-age-optimal policies when multiple hops separate the source and destination.

- *Relaxing the intra-class synchronized sampling and arrivals assumption*: The lex-optimality results provided in Chapter 3 rely on the intra-class synchronized sampling and arrivals assumption. Accordingly, it is compelling to see the effect of this assumption's relaxation on the lex-age-optimality results.

CSMA Environments

In Chapter 4, we have focused on a random access environment where links employ the widely used CSMA scheme. The results of this chapter can be extended by investigating

- *Hidden node problems*: The analysis in Chapter 4 assumed that all transmitters are within the same communication range and, consequently, the problem of hidden nodes does not exist. An interesting direction is to investigate the effect of this problem when it arises on the age performance of the network.

- *Multi-hop*: As it was the case for the multi-class systems, multi-hop scenarios arise naturally in random access environments. Examining multi-hop CSMA random access is compelling since it may shed light on intriguing interactions between the various links. In this case, the priority given to each link in accessing the channel is expected to depend not only on the transmission time but also on the number of hops till destination.

- *General transmission time distributions*: In Chapter 4, we have theoretically investigated the case where the packet transmission time of each link is exponentially distributed. Although the validity of the insights provided in the chapter was verified numerically for various other distributions, an interesting research direction is to validate these insights theoretically.

- *Stochastic arrivals*: We have provided in Chapter 4 an upper bound to the average age of the network, which was used to optimize the system's performance. Although inherently complex, it will be interesting to derive an exact expression of the average age in the stochastic arrival case.

- *General age-based metrics*: In many real-life applications, our interest may lie in minimizing age-based metrics different from the average age. The question on what is the optimal back-off time for each link in the case of a general age-based metric remains open.

Age of Incorrect Information

In Chapter 5, we have shed light on the shortcomings of the age of information and standard error metrics in many real-time applications. To that end, we have introduced the age of incorrect information metric in the status updates framework. The analysis and the model of this chapter can be extended by considering

- *General functions*: In Chapter 5, we have theoretically investigated the case where the time dissatisfaction function $f(t)$ is linear, and the information penalty function $g(X(t), \hat{X}(t))$ is simple indicator function. Our investigations of more general settings revolved solely on numerical implementations. The natural next step in extending the analysis is to derive optimal sampling policies for more general combinations of the functions $f(t)$ and $g(X(t), \hat{X}(t))$.

- *Continuous time models*: In Chapter 5, we examined a discrete-time system. Although the applicability of discrete-time systems in real-life scenarios is broad, one of the potential research directions is to study the AoII metric in continuous-time systems.

- *Real-life applications*: We have provided in Chapter 5 a variety of real-life applications that can be modeled through the proposed AoII metric. As the metric is new, relating it to an even wider range of applications is essential to understand its scope better. Accordingly, an interesting approach is to find the

appropriate combination of both functions $f(t)$ and $g(X(t), \hat{X}(t))$ to model various other real-life applications.

Bibliography

[1] "Internet of things (iot)." https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT. Accessed: February 2020.

[2] J. Manyika, M. Chui, and J. Bughin, "Disruptive technologies: Advances that will transform life, business, and the global economy," 2013.

[3] E. Ahmed and H. Gharavi, "Cooperative vehicular networking: A survey," *IEEE transactions on intelligent transportation systems : a publication of the IEEE Intelligent Transportation Systems Council*, vol. 19, pp. 996–1014, Mar 2018.

[4] 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, pp. 584–616, Fourth 2011.

[5] G. M. Toschi, L. B. Campos, and C. E. Cugnasca, "Home automation networks: A survey," *Computer Standards and Interfaces*, vol. 50, pp. 42 – 54, 2017.

[6] K. Lu, Y. Qian, D. Rodriguez, W. Rivera, and M. Rodriguez, "Wireless sensor networks for environmental monitoring applications: A design framework," in *IEEE GLOBECOM 2007 - IEEE Global Telecommunications Conference*, pp. 1108–1112, 2007.

[7] 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, pp. 7492–7508, Nov 2017.

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

[9] X. Song and J. W. S. Liu, "Performance of multiversion concurrency control algorithms in maintaining temporal consistency," in *Proceedings., Four-*

teenth Annual International Computer Software and Applications Conference, pp. 132–139, 1990.

[10] A. Kosta, N. Pappas, and V. Angelakis, "Age of information : A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, pp. 162–259, 2017.

[11] 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.

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

[13] 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.

[14] Z. Jia, X. Qin, Z. Wang, and B. Liu, "Age-based path planning and data acquisition in uav-assisted iot networks," in *2019 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, 2019.

[15] M. Costa, S. Valentin, and A. Ephremides, "On the age of channel information for a finite-state markov model," in *2015 IEEE International Conference on Communications (ICC)*, pp. 4101–4106, 2015.

[16] M. Bastopcu and S. Ulukus, "Who Should Google Scholar Update More Often?," *arXiv e-prints*, p. arXiv:2001.11500, Jan. 2020.

[17] 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, 2019.

[18] C. Kam, S. Kompella, and A. Ephremides, "Effect of message transmission diversity on status age," in *2014 IEEE International Symposium on Information Theory*, pp. 2411–2415, 2014.

[19] K. Chen and L. Huang, "Age-of-information in the presence of error," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 2579–2583, 2016.

[20] P. Mayekar, P. Parag, and H. Tyagi, "Optimal lossless source codes for timely updates," in *2018 IEEE International Symposium on Information Theory (ISIT)*, pp. 1246–1250, 2018.

[21] J. Zhong and R. D. Yates, "Timeliness in lossless block coding," in *2016 Data Compression Conference (DCC)*, pp. 339–348, 2016.

[22] S. Feng and J. Yang, "Age-optimal transmission of rateless codes in an erasure channel," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, pp. 1–6, 2019.

[23] E. T. Ceran, D. Gündüz, and A. György, "Average age of information with hybrid arq under a resource constraint," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, April 2018.

[24] B. T. Bacinoglu and E. Uysal-Biyikoglu, "Scheduling status updates to minimize age of information with an energy harvesting sensor," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 1122–1126, June 2017.

[25] R. D. Yates, "Lazy is timely: Status updates by an energy harvesting source," in *2015 IEEE International Symposium on Information Theory (ISIT)*, pp. 3008–3012, 2015.

[26] M. Moltafet, M. Leinonen, M. Codreanu, and N. Pappas, "Power Minimization for Age of Information Constrained Dynamic Control in Wireless Sensor Networks," *arXiv e-prints*, p. arXiv:2007.05364, July 2020.

[27] P. Zou, O. Ozel, and S. Subramaniam, "Waiting before serving: A companion to packet management in status update systems," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3864–3877, 2020.

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

[29] J. Xu and N. Gautam, "Towards Assigning Priorities in Queues Using Age of Information," *arXiv e-prints*, p. arXiv:1906.12278, Jun 2019.

[30] 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 Transactions on Networking*, vol. 26, pp. 2637–2650, Dec 2018.

[31] Y. Hsu, E. Modiano, and L. Duan, "Age of information: Design and analysis of optimal scheduling algorithms," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 561–565, June 2017.

[32] Y. 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*, pp. 1–1, 2019.

[33] 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)*, pp. 136–141, April 2018.

[34] A. M. Bedewy, Y. Sun, S. Kompella, and N. B. Shroff, "Age-optimal Sampling and Transmission Scheduling in Multi-Source Systems," *arXiv e-prints*, p. arXiv:1812.09463, Dec 2018.

[35] "A collection of recent papers on the age of information." http://webhome.auburn.edu/~yzs0078/. Accessed: July 2020.

[36] R. Talak, S. Karaman, and E. Modiano, "Distributed Scheduling Algorithms for Optimizing Information Freshness in Wireless Networks," *arXiv e-prints*, p. arXiv:1803.06469, Mar. 2018.

[37] Z. Jiang, B. Krishnamachari, X. Zheng, S. Zhou, and Z. Niu, "Timely Status Update in Massive IoT Systems: Decentralized Scheduling for Wireless Uplinks," *arXiv e-prints*, p. arXiv:1801.03975, Jan. 2018.

[38] Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Can decentralized status update achieve universally near-optimal age-of-information in wireless multiaccess channels?," in *2018 30th International Teletraffic Congress (ITC 30)*, vol. 01, pp. 144–152, 2018.

[39] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *2011 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pp. 350–358, June 2011.

[40] 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*, pp. 1–1, 2019.

[41] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, "Aoi-penalty minimization for networked control systems with packet loss," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 189–196, 2019.

[42] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, pp. 1–5, 2018.

[43] Z. Jiang, S. Zhou, Z. Niu, and Y. Cheng, "A Unified Sampling and Scheduling Approach for Status Update in Multiaccess Wireless Networks," *arXiv e-prints*, p. arXiv:1812.05215, Dec 2018.

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

[45] H. Tang, J. Wang, Z. Tang, and J. Song, "Scheduling to minimize age of synchronization in wireless broadcast networks with random updates," *IEEE Transactions on Wireless Communications*, vol. 19, no. 6, pp. 4023–4037, 2020.

[46] 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)*, pp. 367–372, April 2018.

[47] "IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 11: Wireless LAN Medium Access Control (MAC)and Physical Layer (PHY) Specifications Amendment 5: Enhancements for Higher Throughput," *IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009)*, pp. 1–565, Oct 2009.

[48] 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*, pp. 1–14, 2020.

[49] A. Maatouk, M. Assaad, and A. Ephremides, "On the age of information in a csma environment," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 818–831, 2020.

[50] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "On The Optimality of The Whittle's Index Policy For Minimizing The Age of Information," *arXiv e-prints*, p. arXiv:2001.03096, Jan. 2020.

[51] A. Maatouk, S. E. Hajri, M. Assaad, and H. Sari, "On optimal scheduling for joint spatial division and multiplexing approach in fdd massive mimo," *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 1006–1021, 2019.

[52] A. Maatouk, M. Assaad, and A. Ephremides, "Energy efficient and throughput optimal csma scheme," *IEEE/ACM Transactions on Networking*, vol. 27, no. 1, pp. 316–329, 2019.

[53] A. Maatouk, E. Çalışkan, M. Koca, M. Assaad, G. Gui, and H. Sari, "Frequency-domain noma with two sets of orthogonal signal waveforms," *IEEE Communications Letters*, vol. 22, no. 5, pp. 906–909, 2018.

[54] 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)*, pp. 1–8, 2020.

[55] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "Asymptotically optimal scheduling policy for minimizing the age of information," in *2020 IEEE International Symposium on Information Theory (ISIT)*, pp. 1747–1752, 2020.

[56] 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)*, pp. 1–8, 2019.

[57] 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)*, pp. 102–108, 2019.

[58] 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)*, pp. 325–329, July 2019.

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

[60] E. Caliskan, A. Maatouk, M. Koca, M. Assaad, G. Gui, and H. Sari, "A simple noma scheme with optimum detection," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2018.

[61] J. Denis, A. Maatouk, S. E. Hajri, and M. Assaad, "Stay longer at the network's edge: A novel proactive caching policy through sojourn time," in *2018 IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, 2018.

[62] A. Maatouk, S. E. Hajri, M. Assaad, H. Sari, and S. Sezginer, "Graph theory based approach to users grouping and downlink scheduling in fdd massive mimo," in *2018 IEEE International Conference on Communications (ICC)*, pp. 1–7, 2018.

[63] H. Sari, A. Maatouk, E. Caliskan, M. Assaad, M. Koca, and G. Gui, "On the foundation of noma and its application to 5g cellular networks," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2018.

[64] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1807–1827, 2019.

[65] R. D. Yates, "Age of information in a network of preemptive servers," *CoRR*, vol. abs/1803.07993, 2018.

[66] C. Kam, J. P. Molnar, and S. Kompella, "Age of information for queues in tandem," in *MILCOM 2018 - 2018 IEEE Military Communications Conference (MILCOM)*, pp. 1–6, 2018.

[67] J. Doncel and M. Assaad, "Age of Information in a Decentralized Network of Parallel Queues with Routing and Packets Losses," *arXiv e-prints*, p. arXiv:2002.01696, Feb. 2020.

[68] C. A. Mack, *Stochastic Hybrid Systems*. 2006.

[69] R. D. Yates, "Age of Information in a Network of Preemptive Servers," *ArXiv e-prints*, Mar. 2018.

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

[71] A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal information updates in multihop networks," in *2017 IEEE International Symposium on Information Theory (ISIT)*, pp. 576–580, June 2017.

[72] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *CoRR*, vol. abs/1608.08622, 2016.

[73] 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, pp. 442–452, Jan 1994.

[74] F. Sivrikaya and B. Yener, "Time synchronization in sensor networks: a survey," *IEEE Network*, vol. 18, pp. 45–50, July 2004.

[75] 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.

[76] S. Yun, Y. Yi, J. Shin, and D. Y. Eun, "Optimal csma: A survey," in *2012 IEEE International Conference on Communication Systems (ICCS)*, pp. 199–204, Nov 2012.

[77] L. Jiang and J. Walrand, "A distributed csma algorithm for throughput and utility maximization in wireless networks," *IEEE/ACM Transactions on Networking*, vol. 18, pp. 960–972, June 2010.

[78] L. B. Jiang and S. C. Liew, "Improving throughput and fairness by reducing exposed and hidden nodes in 802.11 networks," *IEEE Transactions on Mobile Computing*, vol. 7, pp. 34–49, Jan 2008.

[79] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications*, vol. 18, pp. 535–547, March 2000.

[80] S. Schaible and J. Shi, "Fractional programming: The sum-of-ratios case," *Optimization Methods and Software*, vol. 18, no. 2, pp. 219–229, 2003.

[81] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1." http://cvxr.com/cvx, Mar. 2014.

[82] R. Yuster and U. Zwick, "Fast sparse matrix multiplication," *ACM Trans. Algorithms*, vol. 1, pp. 2–13, July 2005.

[83] S. Li, *Fast Algorithms for Sparse Matrix Inverse Computations*. PhD thesis, Stanford, CA, USA, 2009. AAI3382775.

[84] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling Policies for Minimizing Age of Information in Broadcast Wireless Networks," *arXiv e-prints*, p. arXiv:1801.01803, Jan 2018.

[85] D. P. Bertsekas, *Dynamic Programming and Optimal Control*. Athena Scientific, 2nd ed., 2000.

[86] E. Altman, *Constrained Markov decision processes*. Chapman and Hall/CRC, 1999.

[87] E. N. Gilbert, "Capacity of a burst-noise channel," *The Bell System Technical Journal*, vol. 39, no. 5, pp. 1253–1265, 1960.

[88] Y. J. Liang, J. G. Apostolopoulos, and B. Girod, "Analysis of packet loss for compressed video: Effect of burst losses and correlation between error frames," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 861–874, 2008.

[89] 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.

[90] M. Razaviyayn, M. Hong, and Z. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[91] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

Appendix A

Multi-Class Multi-Stream Scheduling

A.1 Proof of Theorem 3.3

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 $k = 1, \ldots, I$. Afterward, we show that the PP-MAF-LGFS policy satisfies these rules for all $k = 1, \ldots, I$, 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 exponential distribution's memoryless property, we can obtain the following coupling lemma.

Lemma A.1 ([33, Lemma 1] Stochastic Coupling)**.** *For any given $\mathcal{I}$, 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 we refer to classes 1 till k as the first k classes throughout this proof. Before laying out the proposition, we define the notion of work-conserving policies for the informative packets of a class k.

Definition A.1 (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 A.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 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 A.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 A.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*

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

then

$$(\Delta_{P_1}^{1,[j]})' \leq (\Delta_{\pi_1}^{1,[j]})', \quad j = 1, \ldots, J_1, \tag{A.2}$$

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 A.2. □

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, each stream's age 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 A.2. By induction over time, we obtain

$$\Delta_{P_1}^{1,[j]}(t) \leq \Delta_{\pi_1}^{1,[j]}(t), \quad j = 1, \ldots, J_1, \quad t \geq 0. \tag{A.3}$$

For any symmetric non-decreasing function p_t, and for $t \geq 0$, it holds from (A.3) that

$$\begin{aligned} p_t \circ \mathbf{\Delta}_{P_1}^1(t) =& p_t(\Delta_{P_1}^{1,1}(t), \ldots, \Delta_{P_1}^{1,J_1}(t)) = p_t(\Delta_{P_1}^{1,[1]}(t), \ldots, \Delta_{P_1}^{1,[J_1]}(t)) \\ \leq& p_t(\Delta_{\pi_1}^{1,[1]}(t), \ldots, \Delta_{\pi_1}^{1,[J_1]}(t)) = p_t(\Delta_{\pi_1}^{1,1}(t), \ldots, \Delta_{\pi_1}^{1,J_1}(t)) = p_t \circ \mathbf{\Delta}_{\pi_1}^1(t). \end{aligned} \tag{A.4}$$

By Lemma A.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 (A.4) and Theorem 6.B.30 in [70], we can deduce that

$$[\{p_t \circ \mathbf{\Delta}_P^1(t), t \geq 0\}|\mathcal{I}] \leq_{st} [\{p_t \circ \mathbf{\Delta}_\pi^1(t), t \geq 0\}|\mathcal{I}], \tag{A.5}$$

for all $\mathcal{I}$, $p_t \in \mathcal{P}_{\text{sym}}$ and $\pi \in \Pi_{wc}$. The extension of (A.5) 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. Due to the memoryless property offered by the exponential distribution, letting the server idle before a transmission

leads to the unnecessary staleness of the available packets. A stochastic ordering argument can show this, but the details are omitted. Consequently, (A.5) 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 the unnecessary staleness of the available informative packets. This can be shown by a stochastic ordering argument but the details are omitted. 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 3.6 and (3.42), we have

$$\mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^1_{\pi}(t), t \geq 0\})|\mathcal{I}] = \min_{\pi' \in \Pi} \mathbb{E}[\phi(\{p_t \circ \mathbf{\Delta}^1_{\pi'}(t), t \geq 0\})|\mathcal{I}], \tag{A.6}$$

for all $\mathcal{I}$, $p_t \in \mathcal{P}_{\text{sym}}$ and non-decreasing functional $\phi : \mathbb{V} \mapsto \mathbb{R}$, provided that the expectations in (A.6) 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

$$\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}^1_{P}(t), t \geq 0\})|\mathcal{I}] < \mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}^1_{\pi}(t), t \geq 0\})|\mathcal{I}]. \tag{A.7}$$

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 A.1. From the sufficiency proof, (A.3) holds for our case. In other words,

$$\Delta^{1,[j]}_{P_1}(t) \leq \Delta^{1,[j]}_{\pi_1}(t), \quad j = 1, \ldots, J_1, \quad t \geq 0. \tag{A.8}$$

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

$$p_t \circ \mathbf{\Delta}^1_{P_1}(t) \leq p_t \circ \mathbf{\Delta}^1_{\pi_1}(t). \tag{A.9}$$

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[1], 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,

[1]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.

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 A.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 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 A.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*

$$(\Delta_{P_1}^{1,[j]})' < (\Delta_{\pi_1}^{1,[j]})'. \tag{A.10}$$

Proof. The proof can be found in Appendix A.3. □

Next, to prove (A.7), let us consider the symmetrical non-decreasing penalty function $p' = p_{\text{sum}} \in \mathcal{P}_{\text{sym}}$ and the non-decreasing age penalty functional $\phi_1 = \phi_{avg}$. By taking Lemma A.3 into account, along with (A.8), 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

$$p' \circ \mathbf{\Delta}_{P_1}^1(t) < p' \circ \mathbf{\Delta}_{\pi_1}^1(t), \quad \forall t \in \mathcal{T}. \tag{A.11}$$

By Lemma A.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 (A.9) and (A.11), we obtain:

$$\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}_P^1(t), t \geq 0\}|\mathcal{I})] < \mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}_\pi^1(t), t \geq 0\}|\mathcal{I})]. \tag{A.12}$$

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 is necessary and sufficient for level k lex-age-optimality. In other words, every policy $\pi \in \Pi^k_{\text{lex-opt}}$ 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 A.1. Next, we provide the following Lemma that describes the evolution of the age vector of classes $i = 1, \ldots, k+1$ upon a packet delivery by both π_1 and P_1.

Lemma A.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*

$$\Delta_{P_1}^{i,[j]} = \Delta_{\pi_1}^{i,[j]}, \quad i = 1, \ldots, k, \;\; j = 1, \ldots, J_i, \tag{A.13}$$

$$\Delta_{P_1}^{k+1,[j]} \leq \Delta_{\pi_1}^{k+1,[j]}, \quad j = 1, \ldots, J_{k+1}, \tag{A.14}$$

then

$$(\Delta_{P_1}^{i,[j]})' = (\Delta_{\pi_1}^{i,[j]})', \quad i = 1, \ldots, k, \;\; j = 1, \ldots, J_i, \tag{A.15}$$

$$(\Delta_{P_1}^{k+1,[j]})' \leq (\Delta_{\pi_1}^{k+1,[j]})', \quad j = 1, \ldots, J_{k+1}. \tag{A.16}$$

Proof. The proof can be found in Appendix A.4. □

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 $i = 1, \ldots, k+1$ 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 A.4. By induction over time, we obtain for all $t \geq 0$:

$$\Delta_{P_1}^{i,[j]}(t) = \Delta_{\pi_1}^{i,[j]}(t), \quad i = 1, \ldots, k, \;\; j = 1, \ldots, J_i, \tag{A.17}$$

$$\Delta_{P_1}^{k+1,[j]}(t) \leq \Delta_{\pi_1}^{k+1,[j]}(t), \quad j = 1, \ldots, J_{k+1}. \tag{A.18}$$

For any symmetric non-decreasing function p_t, and for $t \geq 0$, it holds from (A.17) and (A.18)

$$p_t \circ \mathbf{\Delta}_{P_1}^i(t) = p_t \circ \mathbf{\Delta}_{\pi_1}^i(t), \quad i = 1, \ldots, k, \quad t \geq 0, \tag{A.19}$$

$$p_t \circ \mathbf{\Delta}_{P_1}^{k+1}(t) \leq p_t \circ \mathbf{\Delta}_{\pi_1}^{k+1}(t), \quad t \geq 0. \tag{A.20}$$

By Lemma A.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 (A.19)-(A.20) and Theorem 6.B.30 in [70], we can deduce that

$$[\{p_t \circ \mathbf{\Delta}_P^i(t), t \geq 0\}|\mathcal{I}] =_{st} [\{p_t \circ \mathbf{\Delta}_\pi^i(t), t \geq 0\}|\mathcal{I}], \quad i = 1, \ldots, k, \tag{A.21}$$

and

$$[\{p_t \circ \mathbf{\Delta}_P^{k+1}(t), t \geq 0\}|\mathcal{I}] \leq_{st} [\{p_t \circ \mathbf{\Delta}_\pi^{k+1}(t), t \geq 0\}|\mathcal{I}], \tag{A.22}$$

for all $\mathcal{I}$, $p_t \in \mathcal{P}_{\text{sym}}$ and $\pi \in \Pi_{wc} \cap \Pi^k_{\text{lex-opt}}$. The extension of (A.21)-(A.22) 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 idle before a transmission leads to the unnecessary staleness of the packets. A stochastic ordering argument can show this, but the details are omitted. Consequently, (A.21)-(A.22) 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 follows 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 the unnecessary staleness of the available packets. A stochastic ordering argument can show this, but the details are omitted. 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 A.1. From the sufficiency proof for level $k+1$ lex-age-optimality, we have that for all $t \geq 0$:

$$\Delta^{i,[j]}_{P_1}(t) = \Delta^{i,[j]}_{\pi_1}(t), \quad i = 1, \ldots, k, \;\; j = 1, \ldots, J_i, \tag{A.23}$$

$$\Delta^{k+1,[j]}_{P_1}(t) \leq \Delta^{k+1,[j]}_{\pi_1}(t), \quad j = 1, \ldots, J_{k+1}. \tag{A.24}$$

Accordingly, for any symmetric non-decreasing function p_t:

$$p_t \circ \mathbf{\Delta}^{i}_{P_1}(t) = p_t \circ \mathbf{\Delta}^{i}_{\pi_1}(t), \quad i = 1, \ldots, k \quad t \geq 0, \tag{A.25}$$

$$p_t \circ \mathbf{\Delta}^{k+1}_{P_1}(t) \leq p_t \circ \mathbf{\Delta}^{k+1}_{\pi_1}(t), \quad t \geq 0. \tag{A.26}$$

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 A.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

$$\Delta_{P_1}^{k+1,[j]}(t_s^+) < \Delta_{\pi_1}^{k+1,[j]}(t_s^+). \tag{A.27}$$

Afterward, we consider the symmetric non-decreasing penalty function $p' = p_{\text{sum}} \in \mathcal{P}_{\text{sym}}$ and the non-decreasing age penalty functional $\phi_1 = \phi_{avg}$. By taking (A.27) into account, along with (A.23)-(A.24), 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

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

By Lemma A.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 (A.25), (A.26), and (A.28), we obtain:

$$[\{p_t \circ \mathbf{\Delta}_P^i(t), t \geq 0\}|\mathcal{I}] =_{st} [\{p_t \circ \mathbf{\Delta}_\pi^i(t), t \geq 0\}|\mathcal{I}], \quad i = 1,\ldots,k, \tag{A.29}$$

and

$$\mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}_P^{k+1}(t), t \geq 0\}|\mathcal{I})] < \mathbb{E}[\phi_1(\{p' \circ \mathbf{\Delta}_\pi^{k+1}(t), t \geq 0\}|\mathcal{I})]. \tag{A.30}$$

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 3.11, 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.

A.2 Proof of Lemma A.2

Let us denote by $W_j^1(t) = \max\{S_n^{1,j} : A_n^{1,j} \leq t\}$ 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 $W_j^1(t) = W^1(t)$ 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 [33]. For the sake of completeness, we provide a proof of all 3 cases.
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:

$$(\Delta_{P_1}^{1,[j]})' = \Delta_{P_1}^{1,[j]} = t - W^1(t), \quad j = 1,\ldots,J_1. \tag{A.31}$$

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

$$\Delta_{\pi_1}^{1,[j]} \geq (\Delta_{\pi_1}^{1,[j]})' \geq t - W^1(t), \quad j = 1,\ldots,J_1. \tag{A.32}$$

Therefore, (A.2) holds for this case.
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, $(\mathbf{\Delta}^1_{\pi_1})' = \mathbf{\Delta}^1_{\pi_1}$ and (A.2) holds trivially in this scenario.
Case 2: 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:

$$\Delta^{1,j}_{P_1} \geq (\Delta^{1,j}_{P_1})' \geq t - W^1(t), \quad j = 1, \ldots, J_1, \tag{A.33}$$

$$\Delta^{1,j}_{\pi_1} \geq (\Delta^{1,j}_{\pi_1})' \geq t - W^1(t), \quad j = 1, \ldots, J_1. \tag{A.34}$$

We recall that P_1 schedules the stream of class 1 with the highest age. Consequently, the stream of class 1 having the age $\Delta^{1,[1]}_{P_1}$ 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,

$$(\Delta^{1,[J_1]}_{P_1})' = t - W^1(t). \tag{A.35}$$

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 (A.35), we get:

$$(\Delta^{1,[j]}_{P_1})' = \Delta^{1,[j+1]}_{P_1}, \quad j = 1, \ldots, J_1 - 1. \tag{A.36}$$

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

$$(\Delta^{1,[j]}_{\pi_1})' \geq \Delta^{1,[j+1]}_{\pi_1}, \quad j = 1, \ldots, J_1 - 1. \tag{A.37}$$

Combining (A.1), (A.36) and (A.37), we obtain:

$$(\Delta^{1,[j]}_{\pi_1})' \geq \Delta^{1,[j+1]}_{\pi_1} \geq \Delta^{1,[j+1]}_{P_1} = (\Delta^{1,[j]}_{P_1})', \quad j = 1, \ldots, J_1 - 1. \tag{A.38}$$

Also, using (A.34) and (A.35), we can deduce that $(\Delta^{1,[J_i]}_{\pi_1})' \geq t - W^1(t) = (\Delta^{1,[J_i]}_{P_1})'$, which concludes the proof.

A.3 Proof of Lemma A.3

To prove this lemma, we recall that (A.8) 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, (A.10) 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 A.2. *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* $[\boldsymbol{\Delta}^i]$. *In other words, if stream* (i, j) *has a rank* $1 \leq r \leq J_i$, *then:*

- *There exists* $J_i - r$ *streams in the same class having an age that is smaller or equal to* $\Delta^{i,j}$.
- *There exists* $r - 1$ *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 $[\boldsymbol{\Delta}^1_{P_1}]$. On the other hand, let us suppose that the served stream by π_1 has rank $r > 1$ in the sorted age vector $[\boldsymbol{\Delta}^1_{\pi_1}]$. 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 (A.10) 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:

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

$$[\boldsymbol{\Delta}^1_{P_1}](t_s^-) = (10, 9, 8, 1). \tag{A.39}$$

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:

$$[\boldsymbol{\Delta}^1_{\pi'_1}](t_s^+) = (10, 9, 1, 1)$$

$$[\boldsymbol{\Delta}^1_{\pi_1}](t_s^+) = (9, 8, 1, 1) \tag{A.40}$$

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 compared to the stream served by π_1. Consequently, (A.10) holds.

A.4 Proof of Lemma A.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 (A.15). Moreover, as the age vector of class $k+1$ remains unchanged for both policies in this case, (A.16) 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, (A.15) holds naturally. Next, to obtain (A.16), we can proceed similarly to Lemma A.2 for class $k+1$. The details are, therefore, omitted.

Appendix B

Average Age Minimization in a CSMA Environment

B.1 Proof of Theorem 4.2

Similarly to the proof of Theorem 4.1, we start by summarizing the transitions between the discrete states and the resets they induce on the age process $\boldsymbol{x}(t)$ in the following table:

l	$q_l \to q_l'$	$\lambda^{(l)}$	$\boldsymbol{x}\boldsymbol{A}_l$	$\boldsymbol{A}_l$	$\overline{\boldsymbol{v}}_{q_l}\boldsymbol{A}_l$
1	$0 \to 1$	R_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[\overline{v}_{00}, \overline{v}_{01}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$0 \to N$	R_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[\overline{v}_{00}, \overline{v}_{01}]$
$N+1$	$1 \to 0$	H_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[\overline{v}_{10}, \overline{v}_{11}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$N+i$	$i \to 0$	H_i	$[x_1, x_1]$	$\left(\begin{smallmatrix}0&0\\1&1\end{smallmatrix}\right)$	$[\overline{v}_{i1}, \overline{v}_{i1}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2N$	$N \to 0$	H_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$	$[\overline{v}_{N0}, \overline{v}_{N1}]$
$2N+1$	$0 \to 0$	λ_i	$[x_0, 0]$	$\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$	$[\overline{v}_{00}, 0]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$3N+1$	$N \to N$	λ_i	$[x_0, 0]$	$\left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$	$[\overline{v}_{N0}, 0]$

Table B.1: Stochastic arrivals scenario SHS description.

The set of transitions from $l = 1$ till $l = 2N$ are identical to those of the sampling case except for a subtle difference: In transition $l = N + i$, the age process at the

monitor of link i resets to the age of the packet that was delivered x_1. However, as previously explained in the chapter, to avoid tracking the buffer status of link i, we suppose a "fake" update is generated with the same age of the previously transmitted packet x_1. Accordingly, we have $x_1' = x_1$. The final set of transitions from $l = 2N+1$ till $l = 3N + 1$ corresponds to a new packet arrival for link i. This new packet will replace the packet already in the system (either currently being served or waiting to be served). The new packet arrival will keep the same age at the monitor x_0 but will reset the age of the system's packet at link i to 0.

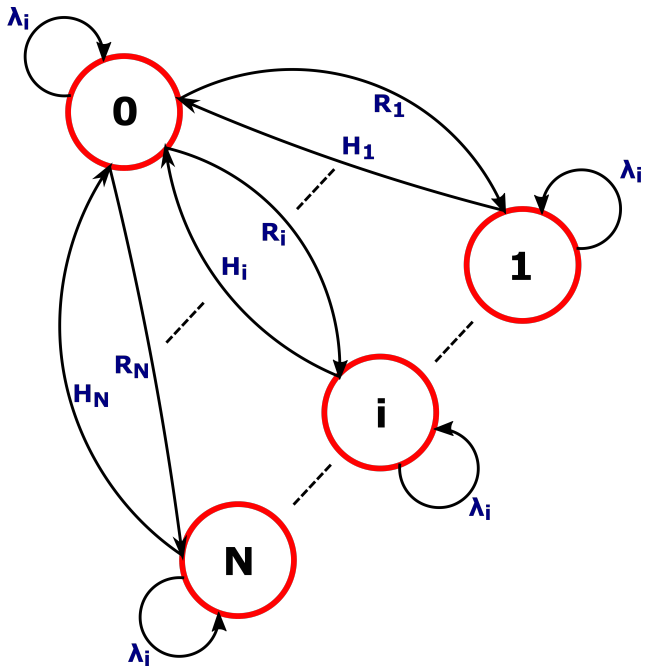


Figure B.1: Illustration of the stochastic hybrid systems Markov chain for the stochastic arrivals scenario.

As for the differential equations governing the evolution of the age process in each discrete state, we have that in each state $q \in \mathbb{Q}$, both $x_0(t)$ and $x_1(t)$ increase at a unit rate:

$$\boldsymbol{b}_q = [1 \;\; 1], \quad \forall q \in \mathbb{Q}. \tag{B.1}$$

It is worth mentioning that the same Markov chain stationary distribution $\overline{\boldsymbol{\pi}}$ of the sampling scenario holds for this case as well. Therefore, by considering the stationary distribution reported in Proposition 4.1, and both the transitions of Table B.1 and the differential equation vector of (B.1), we can solve eqs. (3.6) to find the vectors $\overline{\boldsymbol{v}}_q = [\overline{v}_{q0}, \overline{v}_{q1}]$, $\forall q \in \mathbb{Q}$. By doing so, we end up with the following set of equations:

$$\overline{v}_{00}\Big(\sum_{k=1}^{N} R_k\Big) = \overline{\pi}_0 + \sum_{\substack{k=1 \\ k\neq i}}^{N} H_k \overline{v}_{k0} + H_i \overline{v}_{i1}, \tag{B.2}$$

$$\overline{v}_{01}(\lambda_i + \sum_{k=1}^{N} R_k) = \overline{\pi}_0 + \sum_{\substack{k=1 \\ k\neq i}}^{N} H_k \overline{v}_{k1} + H_i \overline{v}_{i1}, \tag{B.3}$$

$$\overline{v}_{k0}(H_k) = \overline{\pi}_k + R_k \overline{v}_{00}, \qquad k = 1, \ldots, N, \tag{B.4}$$

$$\overline{v}_{k1}(\lambda_i + H_k) = \overline{\pi}_k + R_k \overline{v}_{01}, \qquad k = 1, \ldots, N. \tag{B.5}$$

The first step to solve this set of $2N + 2$ equations consists of calculating $\overline{v}_{01}$. From the N equations presented in (B.5), we can conclude that $\overline{v}_{k1} = \frac{\overline{\pi}_k}{\lambda_i + H_k} + \frac{R_k \overline{v}_{01}}{\lambda_i + H_k}$, $\forall k = 1, \ldots, N$. By replacing these values in (B.3), we can conclude that

$$\overline{v}_{01} = \frac{\overline{\pi}_0 + \sum_{k=1}^{N} \frac{H_k \overline{\pi}_k}{\lambda_i + H_k}}{\lambda_i + \sum_{k=1}^{N} R_k (1 - \frac{H_k}{\lambda_i + H_k})}. \tag{B.6}$$

Knowing that $\overline{v}_{i1} = \frac{\overline{\pi}_i}{\lambda_i + H_i} + \frac{R_i \overline{v}_{01}}{\lambda_i + H_i}$, we can therefore proceed with computing $\overline{v}_{00}$ using (B.2):

$$\overline{v}_{00} = \frac{1 - \overline{\pi}_i + \frac{H_i \overline{\pi}_i}{\lambda_i + H_i}}{R_i} + \frac{(\frac{H_i}{\lambda_i + H_i})(\overline{\pi}_0 + \sum_{k=1}^{N} \frac{H_k \overline{\pi}_k}{\lambda_i + H_k})}{\lambda_i + \sum_{k=1}^{N} R_k (1 - \frac{H_k}{\lambda_i + H_k})}. \tag{B.7}$$

The next step consists of using the results of (4.4). By replacing $\overline{\pi}_k$ $0 \leq k \leq N$ with their values and after some algebraic manipulations, we end up with

$$\overline{v}_{00} = \frac{1}{R_i} + \frac{1}{C(\boldsymbol{R})} (\frac{1}{\lambda_i} - \frac{1}{H_i}). \tag{B.8}$$

Next, we know that[1] $\overline{\Delta}_i(\boldsymbol{R}) = \overline{v}_{00} + \sum_{k=1}^{N} \overline{v}_{k0}$. By using the N equations of (B.4), we get that $\sum_{k=1}^{N} \overline{v}_{k0} = \sum_{k=1}^{N} \frac{\overline{\pi}_k}{H_k} + \overline{v}_{00}(C(\boldsymbol{R}) - 1)$, and the average age of the link of interest i is therefore

$$\overline{\Delta}_i(\boldsymbol{R}) = \frac{C(\boldsymbol{R})}{R_i} + \frac{1}{\lambda_i} - \frac{1}{H_i} + \frac{\sum_{k=1}^{N} \frac{R_k}{H_k^2}}{C(\boldsymbol{R})}. \tag{B.9}$$

These results are general and hold for any link i in the network. Therefore, by summing over all the links i, the total average age can be concluded.

[1]The non-negativity of $\overline{\boldsymbol{v}}$ can be easily verified by replacing $C(\boldsymbol{R})$ with its value from eq. (4.5). Consequently, the SHS is stable.

B.2 Proof of Theorem 4.3

First of all, we can see that the formulated problem in (4.22) is convex in $\boldsymbol{f} = [f_1, \ldots, f_N]$ and ϵ as the objective function is the sum of convex functions in $(\boldsymbol{f}, \epsilon)$ and the constraints are linear. We therefore formulate the Lagrange function as follows:

$$\begin{aligned}\mathcal{L}(\boldsymbol{f}, \epsilon, \nu, \gamma, \rho, \boldsymbol{\mu}, \boldsymbol{\eta}) =& N \sum_{k=1}^{N} \frac{f_k}{H_k^2} + \sum_{k=1}^{N} \frac{1}{f_k} + \nu(\epsilon - 1) + \gamma\left(\frac{1}{1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}} - \epsilon\right) \\ &+ \sum_{k=1}^{N} \mu_k (f_k - \epsilon R_{UB}) + \sum_{k=1}^{N} \eta_k (-f_k) + \rho\left(\sum_{k=1}^{N} \frac{f_k}{H_k} + \epsilon - 1\right).\end{aligned} \tag{B.10}$$

As the problem was shown to be convex, we formulate the Karush-Kuhn-Tucker (KKT) conditions which are sufficient for optimality in our case:

$$\frac{N}{H_k^2} - \frac{1}{f_k^{*2}} + \mu_k^* - \eta_k^* + \frac{\rho^*}{H_k} = 0, \qquad k = 1, \ldots, N, \tag{B.11}$$

$$\rho^* + \nu^* - \gamma^* - R_{UB} \sum_{k=1}^{N} \mu_k^* = 0, \tag{B.12}$$

$$\gamma^* \left(\frac{1}{1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}} - \epsilon^*\right) = 0, \tag{B.13}$$

$$\nu^* (\epsilon^* - 1) = 0, \tag{B.14}$$

$$\mu_k^* (f_k^* - \epsilon^* R_{UB}) = 0, \qquad k = 1, \ldots, N, \tag{B.15}$$

$$\eta_k^* f_k^* = 0, \qquad k = 1, \ldots, N, \tag{B.16}$$

$$\nu^*, \gamma^* \geq 0, \qquad \boldsymbol{\mu}, \boldsymbol{\eta} \geq \mathbf{0}, \tag{B.17}$$

$$0 \leq f_k^* \leq \epsilon^* R_{UB}, \qquad k = 1, \ldots, N, \tag{B.18}$$

$$\frac{1}{1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}} \leq \epsilon^* \leq 1, \tag{B.19}$$

$$\sum_{k=1}^{N} \frac{f_k^*}{H_k} = 1 - \epsilon^*. \tag{B.20}$$

In the following, we find the optimal solution using the above *sufficient* optimality conditions. First we suppose that $\nu^* > 0$, which means that $\epsilon^* = 1$. Replacing this in eq. (B.20), and knowing that $f_k^* \geq 0, \ \forall k$, we get that $f_k^* = 0, \ \forall k$, and the objective

function tends to $+\infty$ which is surely not optimal. Therefore, we can conclude that $\nu^* = 0$. The same argument can be used to show that $\eta_k = 0$, $\forall k$. We now suppose that $\gamma^* > 0$, which entails that $\epsilon^* = \frac{1}{1+\sum_{k=1}^{N} \frac{R_{UB}}{H_k}}$. By replacing this in eq. (B.20), and by noting the conditions in eq. (B.18), we can conclude that this is only feasible when $f_k^* = \epsilon^* R_{UB}$, $\forall k$. We replace f_k^* and ϵ^* by their values in eq. (B.11) and we end up with:

$$\mu_k^* = -\frac{\rho^*}{H_k} + \frac{\left(1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}\right)^2}{R_{UB}^2} - \frac{N}{H_k^2}, \quad k = 1, \ldots, N. \tag{B.21}$$

Using the fact that $\mu_k^* \geq 0$, we have the following N conditions on ρ^*:

$$\rho^* \leq \frac{H_k\left(1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}\right)^2}{R_{UB}^2} - \frac{N}{H_k}, \quad k = 1, \ldots, N. \tag{B.22}$$

Knowing that $\mu_k \geq 0$, and by using eq. (B.12), we can conclude that $\rho^* \geq \gamma^* > 0$. To proceed with this case, we define $x = \sum_{k=1}^{N} \frac{1}{H_k}$ and $y = \sum_{k=1}^{N} \frac{1}{H_k^2}$. By summing eqs. (B.11) for all k and using the results of eq. (B.12), we end up with:

$$\gamma^* = NyR_{UB} - \frac{N\left(1 + \sum_{k=1}^{N} \frac{R_{UB}}{H_k}\right)^2}{R_{UB}} + \rho^*(1 + xR_{UB}). \tag{B.23}$$

As $\gamma^* > 0$, the final condition on ρ^* is therefore

$$\rho^* > \frac{N(1 + xR_{UB})}{R_{UB}} - \frac{NyR_{UB}}{1 + xR_{UB}}. \tag{B.24}$$

If $\exists \rho^*$ such that the conditions reported in eqs. (B.22) and (B.24) are verified then $f_k^* = \epsilon^* R_{UB}$, $\forall k$ and the original problem's optimal point is $R_k = R_{UB}$, $\forall k$. We will show that this is always achieved when $H_k = H$, $\forall k$ in the next Lemma. In the latter cases where $\gamma^* = 0$, this entails that there could be at least one link k such that $f_k^* < \epsilon^* R_{UB}$. Therefore, in this scenario, the optimal solution f_k^*, $\forall k$ is such that

$$f_k^* = \begin{cases} \epsilon^* R_{UB}, & \text{if } \mu_k^* > 0, \\ \sqrt{\frac{H_k}{\frac{N}{H_k} + \rho^*}}, & \text{if } \mu_k^* = 0, \end{cases} \tag{B.25}$$

where μ_k^*, ρ^* and ϵ^* verify:

$$\mu_k^* = -\frac{N}{H_k^2} + \frac{1}{f_k^{*2}} - \frac{\rho^*}{H_k}, \qquad k = 1, \ldots, N, \tag{B.26}$$

$$\rho^* = R_{UB} \sum_{k=1}^{N} \mu_k^*,$$
$$\epsilon^* = 1 - \sum_{k=1}^{N} \frac{f_k^*}{H_k}. \tag{B.27}$$

The optimal solution $(\boldsymbol{f}^*, \epsilon^*)$ of the problem in (4.22) can therefore be found. The optimal back-off rate of each link k can then be deduced by noting that $R_k^* = \frac{f_k^*}{\epsilon^*}$, $\forall k$.

B.3 Proof of Lemma 4.1

To prove this lemma, it is sufficient to show that the conditions on ρ^* of eqs. (B.22) and (B.24) are always verified for all feasible (N, H, R_{UB}). More specifically, we have to show that

$$\frac{H\left(1 + \frac{NR_{UB}}{H}\right)^2}{R_{UB}^2} - \frac{N}{H} > \frac{N(1 + \frac{NR_{UB}}{H})}{R_{UB}} - \frac{\frac{N^2 R_{UB}}{H^2}}{1 + \frac{NR_{UB}}{H}}. \tag{B.28}$$

By taking a common denominator, and knowing that $H, R_{UB} > 0$, the positivity of the expression above is equivalent to that of the following cubic polynomial $f(N, H, R_{UB})$:

$$(H + NR_{UB})^3 - NR(H + NR_{UB})^2 - NHR_{UB}^2 > 0. \tag{B.29}$$

By deriving $f(N, H, R_{UB})$ with respect to H, we find that:

$$\frac{\partial f(N, H, R_{UB})}{\partial H} = 3H^2 + H(4NR_{UB}) + N^2 R_{UB}^2 - NR_{UB}^2. \tag{B.30}$$

To study the sign of $\frac{\partial f(N,H,R_{UB})}{\partial H}$, we derive one more time with respect to H and we can show that

$$\frac{\partial^2 f(N, H, R_{UB})}{\partial H^2} = 6H + 4NR_{UB} > 0. \tag{B.31}$$

Therefore, $\frac{\partial f(N,H,R_{UB})}{\partial H}$ is always increasing, bearing in mind that

$$\lim_{H \to 0} \frac{\partial f(N, H, R_{UB})}{\partial H} = N^2 R_{UB}^2 - NR_{UB}^2 \geq 0, \tag{B.32}$$

since $N \geq 1$. Hence, we can assert that $\frac{\partial f(N,H,R_{UB})}{\partial H} > 0$ and, consequently, that $f(N, H, R_{UB})$ increases with H. By following the same approach, we can find that $\frac{\partial f(N,H,R_{UB})}{\partial R_{UB}} = 2NH^2 + 2N^2 HR_{UB} - 2NHR_{UB}$, and $\frac{\partial^2 f(N,H,R_{UB})}{\partial R_{UB}^2} = 2NH(N - 1) \geq 0$ since $N \geq 1$. As $\lim_{R_{UB} \to 0} \frac{\partial f(N,H,R_{UB})}{\partial R_{UB}} = 2NH^2 > 0$, we can conclude that $f(N, H, R_{UB})$ increases with R_{UB}. We can use the same argument over

N by relaxing its discrete nature to a continuous one $n \in [1, +\infty[$, and by noting that $\frac{\partial f(n,H,R_{UB})}{\partial n} = 2R_{UB}H^2 + 2nR_{UB}^2H - R_{UB}^2H$, $\frac{\partial^2 f(n,H,R_{UB})}{\partial n^2} = 2R_{UB}^2H$ and $\frac{\partial f(n,H,R_{UB})}{\partial n}\Big|_{n=1} = 2R_{UB}H^2 + R_{UB}^2H$, we can show that $f(N, H, R_{UB})$ increases with N. Knowing that $f(1, H, R_{UB})$ increases with (H, R_{UB}) ,and that $\lim_{(R_{UB},H)\to(0,0)} f(1, H, R_{UB}) \longrightarrow 0$, we can therefore assert that $\forall\delta_1, \delta_2 > 0$ such that $H \geq \delta_1, R_{UB} \geq \delta_2, \exists\delta > 0$ such that $f(N, H, R_{UB}) \geq \delta$ with $\delta = (\delta_1 + \delta_2)^3 - \delta_2(\delta_1 + \delta_2)^2 - \delta_2^2(\delta_1 + \delta_2) + \delta_2^3 > 0$. This concludes our proof.

B.4 Proof of Theorem 4.4

To proceed with the proof, we investigate the network from the perspective of link i. Therefore, the process $\boldsymbol{x}(t)$ is defined as $\boldsymbol{x}(t) = [x_0(t), x_1(t)]$, where $x_0(t)$ is the age of link i at the monitor and $x_1(t)$ is the age of the packet in the system of link i. Our goal is to solve eqs. (3.6) to conclude the vectors $\overline{\boldsymbol{v}}_q = [\overline{v}_{q0}, \overline{v}_{q1}]$, $\forall q \in \mathbb{Q}$ that will allow us to calculate the average age through Theorem 3.1. We report, in Table B.2, the reset maps that the transitions depicted in the Markov chain of Fig. 4.7 induce on the continuous process. The details are omitted as the logic is similar to that reported in earlier sections. It is worth noting that the additional transitions of rate w correspond to link i switching to AWAKE mode.

l	$q_l \to q_l'$	$\lambda^{(l)}$	$\boldsymbol{x}\boldsymbol{A}_l$	$\boldsymbol{A}_l$	$\overline{\boldsymbol{v}}_{q_l}\boldsymbol{A}_l$
1	$0 \to 1$	R_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{00}, \overline{v}_{01}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	$0 \to N$	R_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{00}, \overline{v}_{01}]$
$N+1$	$1 \to 0$	H_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{10}, \overline{v}_{11}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$N+i$	$i \to 0'$	H_i	$[x_1, 0]$	$\left(\begin{smallmatrix}0 & 0\\ 1 & 0\end{smallmatrix}\right)$	$[\overline{v}_{i1}, 0]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$2N$	$N \to 0$	H_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{N0}, \overline{v}_{N1}]$
$2N+1$	$0' \to 1'$	R_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{00}, \overline{v}_{01}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$3N-1$	$0' \to N'$	R_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{0'0}, \overline{v}_{0'1}]$
$3N$	$1' \to 0'$	H_1	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{1'0}, \overline{v}_{1'1}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$4N-2$	$N' \to 0'$	H_N	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{N'0}, \overline{v}_{N'1}]$
$4N-1$	$0' \to 0$	w	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{0'0}, \overline{v}_{0'1}]$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$5N-2$	$N' \to N$	w	$[x_0, x_1]$	$\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$	$[\overline{v}_{N'0}, \overline{v}_{N'1}]$

Table B.2: SHS description from the perspective of link i.

As for the differential equations governing the evolution of the age process in each discrete state, we know that the age at the monitor increases at a unit rate $\forall q \in \mathbb{Q}$. On the other hand, link i samples the process when it captures the channel, and therefore $x_1(t)$ increases at a unit rate when $q(t) = i$. Therefore, we have that $\boldsymbol{b}_q = [1 \;\; 0]$, $\forall q \neq i$ and $\boldsymbol{b}_i = [1 \;\; 1]$. We can now apply Theorem 3.1 to find the average age of link i. As a first step, it is easy to verify that $\overline{v}_{q1} = 0$, $\forall q \neq i$, as in these states there is no packet in the system for link i. We then start by solving eqs. (3.6), more specifically at state k' which leads to eq. (4.38). Similarly, the results of eq. (4.40) can be obtained by solving eqs. (3.6) in state i. By formulating eqs. (3.6) in state $q = 0'$, and by taking into account the results of eq. (4.38), and the fact that $\overline{v}_{i1} = \frac{\overline{\pi}_i}{H_i}$, the results of eq. (4.37) can be deduced. Finally, by combining the results of eqs. (3.6) at states 0 and $k \neq i$, and by taking into account the previous mentioned results of eqs. (4.37)-(4.38), the results of eq. (4.39) and (4.41) can be obtained. Note that the non-negativity of $\overline{\boldsymbol{v}}$ is straightforward. Accordingly, we can then assert that the SHS is stable. This concludes our proof.

B.5 Proof of Theorem 4.5

In the sequel, we look at the network from the perspective of link $j \neq i$. Similarly, we define the continuous-time state process as $\boldsymbol{x}(t) = [x_0(t), x_1(t)]$, where $x_0(t)$ is the age of link j at the monitor at time t and $x_1(t)$ is the age of the packet in the system of link j at time t. The same transitions depicted in the Markov chain of Fig. 4.7 still hold for the system, but what differs are the resets they induce to $\boldsymbol{x}(t)$. More specifically, the same resets of Table B.2 hold except the following: 1) the transition of rate H_i does not induce a reset, i.e., $\boldsymbol{x'} = \boldsymbol{x}\boldsymbol{A}_l = \boldsymbol{x}$, and 2) the transitions of rate H_j originating from state j' and j will both reset the age to $\boldsymbol{x'} = [x_1, 0]$. In regards to the differential equations governing the evolution of the age process in each discrete state, we know that the age at the monitor increases at a unit rate $\forall q \in \mathbb{Q}$. On the other hand, link j samples the process when it captures the channel, and therefore $x_1(t)$ increases at a unit rate when $q(t) = j$ or $q(t) = j'$:

$$\boldsymbol{b}_q = [1 \;\; 0], \quad \forall q \neq j, j', \tag{B.33}$$

$$\boldsymbol{b}_q = [1 \;\; 1], \quad q = j, j'. \tag{B.34}$$

We now proceed with applying Theorem 3.1 to find a closed-form expression of the average age of link j. Similarly to the previous theorem, and for the same reasons, one can easily verify that $\overline{v}_{q1} = 0, \quad \forall q \neq j, j'$. We then solve eqs. (3.6) at state i and states $k', \ \forall k'$ to get the results of eq. (4.45) and (4.44). Next, by investigating the same equations (3.6) at state $0'$, and by using the results of eq. (4.44) and (4.45), and by noting that $\overline{v}_{j'1} = \frac{\overline{\pi}_{j'}}{H_j + w}$, the results of eq. (4.43) can be found. Eq. (4.46) can be deduced through equations (3.6) in states $k \neq i$. Lastly, by combining all the aforementioned results, and by solving equations (3.6) in state 0, $\overline{v}_{00}$ can be calculated. It is worth noting that the non-negativity of $\overline{\boldsymbol{v}}$ can be easily proven by further simplifying the expression in the denominator of eq. (4.42). Accordingly, the SHS in question is stable. This concludes the theorem.

B.6 Proof of Proposition 4.3

The proof revolves around showcasing different key characteristics of the SCA function $\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$ in (4.52) and leveraging them to demonstrate the desired results. The characteristics are summarized in the following:

- $\overline{\Delta}(\boldsymbol{x}) \leq \overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y}), \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$
- $\lim_{\boldsymbol{x} \to \boldsymbol{y}} \nabla_{\vec{\boldsymbol{d}}} \overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y}) = \nabla_{\vec{\boldsymbol{d}}} \overline{\Delta}(\boldsymbol{y}), \quad \forall \vec{\boldsymbol{d}}, \boldsymbol{y} : \boldsymbol{y}, \vec{\boldsymbol{d}} + \boldsymbol{y} \in \mathcal{X}$
- $\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$ is continuous $\forall (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{X} \times \mathcal{X}$
- $\overline{\Upsilon}(\boldsymbol{y}, \boldsymbol{y}) = \overline{\Delta}(\boldsymbol{y}), \quad \forall \boldsymbol{y} \in \mathcal{X}$

When the function $\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$ has the above characteristics, and by taking into account that $\mathcal{X}$ is a compact set, we can assert that the assumptions in [90, Assumption 1] are all met. We can, therefore, apply [90, Theorem 1] to prove that every limit point of the iterates generated by the algorithm in (4.51) is a stationary point of the problem in (4.50).
We start our proof by tackling the first characteristic of the function $\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$. To do so, we first recall the expression of the objective function:

$$\overline{\Delta}(\boldsymbol{x}) = \sum_{m=1}^{N} \boldsymbol{c}\boldsymbol{E}_m^{-1}\boldsymbol{F}_m\boldsymbol{A}^{-1}\boldsymbol{d}. \tag{B.35}$$

The ergodicity of the Markov chain was previously established for any $\boldsymbol{x} \in \mathcal{X}$, which leads to the existence of $\boldsymbol{A}^{-1}$ (i.e., $\det(\boldsymbol{A}) \neq 0$). Therefore, we have that:

$$\left(\boldsymbol{A}^{-1}\right)_{ij} = \frac{(-1)^{i+j} M_{ji}}{\det(\boldsymbol{A})}, \tag{B.36}$$

where M_{ji} is the (j, i) minor of $\boldsymbol{A}$. By investigating the equations reported in (4.47), we can conclude that the entries $\left(\boldsymbol{A}\right)_{ij}$ are simply linear functions of the components of $\boldsymbol{x}$. Consequently, the entries $\left(\boldsymbol{A}^{-1}\right)_{ij}$ are all rational functions of $\boldsymbol{x}$ with the denominator being $\det(\boldsymbol{A})$. Similarly, one can show that the entries $\left(\boldsymbol{E}_m^{-1}\right)_{ij}$, $\forall m$ are all rational functions of $\boldsymbol{x}$ with the denominator being $\det(\boldsymbol{E}_m)$. As it was previously mentioned, we have that $\det(\boldsymbol{E}_m) \neq 0$ for any $\boldsymbol{x} \in \mathcal{X}$. Moreover, as the matrices $\boldsymbol{F}_m$, $\forall m$ and vectors $\boldsymbol{c}$,$\boldsymbol{d}$ are constant with entries' values equal to 0 and 1, one can deduce that the overall age function $\overline{\Delta}(\boldsymbol{x})$ can be written as a rational function of $\boldsymbol{x}$ with non-zero denominator for any $\boldsymbol{x} \in \mathcal{X}$. Consequently, we can assert that $\overline{\Delta}(\boldsymbol{x})$ is a continuous and differentiable function on the set $\mathcal{X}$. The same argument can be made for the gradient function $\nabla\overline{\Delta}(\boldsymbol{x})$ and, accordingly, the entries of the Hessian matrix $\nabla^2\overline{\Delta}(\boldsymbol{x})$. Due to the continuity of the Hessian matrix, and as the set $\mathcal{X}$ is compact, we know that there exists a constant $L > 0$ such that $||\nabla^2\overline{\Delta}(\boldsymbol{x})|| \leq L$. By using the mean value theorem on the function $\nabla\overline{\Delta}(\boldsymbol{x})$, and by noting the bound on $||\nabla^2\overline{\Delta}(\boldsymbol{x})||$, we can show that $\nabla\overline{\Delta}(\boldsymbol{x})$ is a Lipschitz function. More specifically, there exist a constant $L > 0$ such that:

$$||\nabla\overline{\Delta}(\boldsymbol{x}) - \nabla\overline{\Delta}(\boldsymbol{y})|| \leq L||\boldsymbol{x} - \boldsymbol{y}||, \quad \forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}. \tag{B.37}$$

Using the results of eq. (B.37), we can apply the descent Lemma [91, Proposition A.24] to show that $\overline{\Delta}(\boldsymbol{x}) \leq \overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$, $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ if $\alpha_n \geq \frac{L}{2}$.
The second characteristic we tackle revolves around the directional derivative of the function $\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$ with respect to $\boldsymbol{x}$. More specifically:

$$\nabla_{\vec{\boldsymbol{d}}}\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y}) = \lim_{\lambda \to 0} \frac{\overline{\Upsilon}(\boldsymbol{x} + \lambda\vec{\boldsymbol{d}}, \boldsymbol{y}) - \overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})}{\lambda}. \tag{B.38}$$

By replacing $\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$ with its value from eq. (4.52), we can show that $\lim_{\boldsymbol{x} \to \boldsymbol{y}} \nabla_{\vec{\boldsymbol{d}}}\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y}) = \langle\nabla\overline{\Delta}(\boldsymbol{y})^T, \vec{\boldsymbol{d}}\rangle = \nabla_{\vec{\boldsymbol{d}}}\overline{\Delta}(\boldsymbol{y})$, where $\langle\cdot,\cdot\rangle$ denotes the dot product.

As for the third characteristic, one can verify that it holds by noting the expression of $\overline{\Upsilon}(\boldsymbol{x}, \boldsymbol{y})$ in (4.52) and by taking into account that $\overline{\Delta}(\boldsymbol{x})$ and $\nabla\overline{\Delta}(\boldsymbol{x})$ were shown to be continuous $\forall \boldsymbol{x} \in \mathcal{X}$. Lastly, by simple substitution, we can confirm that $\overline{\Upsilon}(\boldsymbol{y}, \boldsymbol{y}) = \overline{\Delta}(\boldsymbol{y})$, $\forall \boldsymbol{y} \in \mathcal{X}$. As all the required assumptions hold, we can assert that the limit point of the sequence $\left\{\overline{\boldsymbol{x}}[n]\right\}_{n=1}^{+\infty}$ is a stationary point of the problem in (4.50) if $\alpha_n \geq \frac{L}{2}$, $n \in \mathbb{N}$.

Appendix C

Age of Incorrect Information: Analysis and Optimization

C.1 Proof of Lemma 5.1

Our proof is based on the well-known Value Iteration Algorithm (VIA) [85]. By letting $J_t(.)$ be the value function at iteration t, the VIA consists of updating the value function as follows:

$$J_{t+1}(S) = \min_{\psi \in \{0,1\}} \Big\{ S + \sum_{S' \in \mathbb{N}} \Pr(S \rightarrow S' | \psi) J_t(S') \Big\}, \quad \forall S \in \mathbb{N}. \tag{C.1}$$

Regardless of the initial value $J_0(S)$, it is well-known that the algorithm converges to the value function of the Bellman equation (5.15) [85] (i.e., $\lim_{t \rightarrow +\infty} J_t(S) = J(S), \ \forall S \in \mathbb{N}$). Consequently, to infer on the monotonicity of $J(S)$, it is sufficient to prove that $\forall S_2 \geq S_1$:

$$J_t(S_2) \geq J_t(S_1), \quad t = 0, 1, \ldots \tag{C.2}$$

To proceed in that direction, and without loss of generality, we suppose that $J_0(S) = 0, \ \forall S \in \mathbb{N}$. Therefore, (C.2) holds for $t = 0$. Next, we suppose that the condition in (C.2) is true up till $t > 0$ and we examine if it holds for $t + 1$. To do so, we examine the Right Hand Side (RHS) of (C.1) for both states S_2 and S_1. To that extent, we first take the case where $S_1 \neq 0$ and we distinguish between the two possible transmission decisions ψ:

- $\psi = 0$: In this case, the RHS is equal to $x = S_1 + (p_R + (N-2)p_t) J_t(S_1 + 1) + p_t J_t(0)$ and $y = S_2 + (p_R + (N-2)p_t) J_t(S_2 + 1) + p_t J_t(0)$ for S_1 and S_2 respectively. Baring in mind that $J_t(S_2) \geq J_t(S_1)$, we can easily see that $x \leq y$.

- $\psi = 1$: In this case, the RHS is equal to $z = S_1 + (p_R p_f + (N-2)p_t + p_s p_t) J_t(S_1 + 1) + (p_R p_s + p_f p_t) J_t(0)$ and $w = S_2 + (p_R p_f + (N-2)p_t + p_s p_t)$

$J_t(S_2+1)+(p_R p_s+p_f p_t)J_t(0)$ for S_1 and S_2 respectively. Taking into account that $J_t(S_2) \geq J_t(S_1)$, we can also verify that $z \leq w$.

Lastly, we know that if $x \leq y$ and $z \leq w$ then $\min(x,z) \leq \min(y,w)$. For the case where $S_1 = 0$, we can show that $x = z = p_R J_t(0) + (N-1)p_t J_t(1)$. After some algebraic manipulations, we can easily verify that the same above inequalities still hold. Consequently, we can assert that $J_{t+1}(S_2) \geq J_{t+1}(S_1), \ \forall t, S_1, S_2 \in \mathbb{N}$. This concludes our inductive proof that shows that the function $J(S)$ is increasing in $S, \ \forall S \in \mathbb{N}$.

C.2 Proof of Theorem 5.1

As we have previously stated, it is well-known that the optimal transmission policy can be obtained by solving the Bellman equation in (5.15). We also recall that the VIA converges to the Bellman equation's value function in (5.15). Consequently, we can deduce the optimal sequence of actions based on the value function at each time instant t by reconsidering the VIA:

$$J_{t+1}(S) = \min_{\psi \in \{0,1\}} \Big\{ S + \sum_{S' \in \mathbb{N}} \Pr(S \rightarrow S' | \psi) J_t(S') \Big\}, \quad \forall S \in \mathbb{N}. \tag{C.3}$$

To that extent, let us define $\Delta J_{t+1}(S)$ as the difference between the value functions if the transmitter sends a packet or remains idle for any state S. More specifically, we have that $\Delta J_{t+1}(S) = J^1_{t+1}(S) - J^0_{t+1}(S)$ where $J^1_{t+1}(S)$ and $J^0_{t+1}(S)$ are the value functions at time $t+1$ if $\psi(t) = 1$ and $\psi(t) = 0$ respectively. By obeying to the dynamics reported in Section 5.3.2, we have:

$$\Delta J_{t+1}(0) = 0, \tag{C.4}$$

$$\Delta J_{t+1}(S) = p_s(p_t - p_R)(J_t(S+1) - J_t(0)), \quad \forall S \in \mathbb{N}^*. \tag{C.5}$$

The first thing we see is that when the system's state is $S(t) = 0$, both actions of remaining idle or transmitting leads to the same value function at time $t+1$. We can now tackle the case where $S(t) \neq 0$. To that extent, and as $J_t(S)$ is always increasing with S (Lemma 5.1), we can assert that $(J_t(S+1) - J_t(0)) \geq 0$. Based on this, we distinguish between the following cases:

Case 1 - $p_t < p_R$: In this scenario, we can see that $\Delta J_{t+1}(S)$ is always negative for any $S \neq 0$. Consequently, it is always optimal to transmit a packet when $S(t) \neq 0$. Combined with the fact that a transmission or remaining idle leads to the same value function when $S(t) = 0$, we can conclude that the optimal policy is to either send updates at each time slot or send updates when the receiver is in an erroneous state (i.e., when $S(t) \neq 0$). To calculate the average cost, we can see that in the case of an "always update" policy, the MDP can be modeled through a DTMC where:

- The states refer to the values of the penalty function $S(t)$.

- The dynamics of $S(t)$, $\forall(S,t)$ coincide with those of $\psi(t) = 1$ of Section 5.3.2.

The DTMC mentioned above is reported in Fig. C.1.

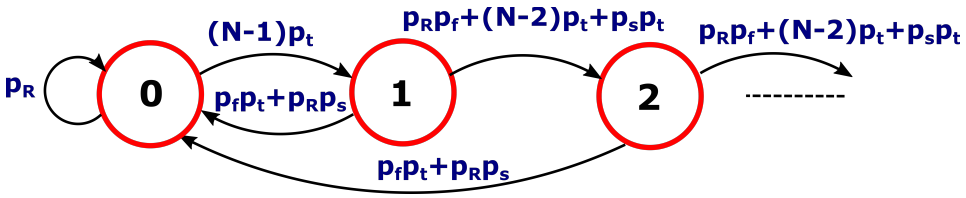


Figure C.1: The states transitions under the “always update" policy.

To find the average cost in this case, we first provide the following lemma.

Lemma C.1. *The DTMC of the “always update" policy is irreducible and admits* π_k, $\forall k \in \mathbb{N}$ *as its stationary distribution where:*

$$\pi_0 = \frac{1}{1 + \frac{(N-1)p_t}{1-a}}, \tag{C.6}$$

$$\pi_k = (N-1)p_t a^{k-1}\pi_0, \quad k \geq 1, \tag{C.7}$$

with the constant a *being equal to* $p_R p_f + (N-2)p_t + p_s p_t$.

Proof. It is sufficient to formulate the general balance equations at any state $k \geq 2$, which leads to $\pi_k = a\pi_{k-1}$. By proceeding with a forward induction, and knowing that $\pi_1 = (N-1)p_t\pi_0$, the results of (C.7) can be found. Next, by taking into account the fundamental equality $\sum_{k=0}^{+\infty} \pi_k = 1$, we can find π_0. □

To find the above DTMC's average cost, we first note that the cost incurred by being at state $S = k$ is nothing but the value k of the state itself. Consequently, we have that $\overline{C}_{AU} = \sum_{k=1}^{+\infty} k\pi_k$. By taking into account the above stationary distribution and the following series equalities, the expression in (5.16) can be found.

$$\sum_{k=1}^{+\infty} a^{k-1} = \frac{1}{1-a}, \qquad \sum_{k=1}^{+\infty} k a^{k-1} = \frac{1}{(1-a)^2}. \tag{C.8}$$

Case 2 **-** $p_t \geq p_R$: In this case, we can see that $\Delta J_{t+1}(S)$ is always positive. Combined with the fact that a transmission or remaining idle leads to the same value function when $S(t) = 0$, we can conclude that the optimal policy is always to remain idle. The intuition behind this is that when $p_t \geq p_R$, any packet being transmitted about the information source has a high chance of becoming obsolete by the time it reaches the monitor. To calculate the average cost in the case where the transmitter is always idle, the MDP can be modeled through the DTMC reported in Fig. C.2. The analysis of this DTMC is the same as the one of the previous case ($p_t < p_R$). More specifically, it is sufficient to substitute a by b where $b = p_R + (N-2)p_t$ in (5.16) to obtain the expression in (5.17).

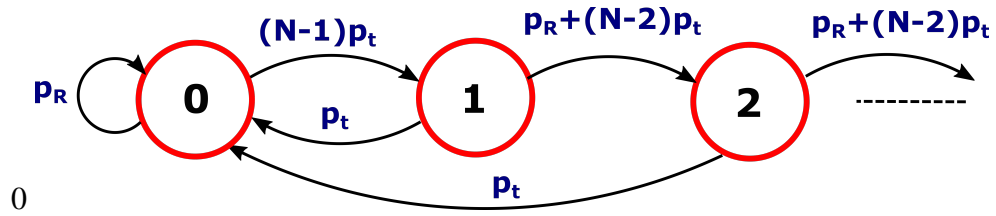


Figure C.2: The states transitions under the “never transmit" policy.

C.3 Proof of Proposition 5.1

The proof follows the same direction as that of Theorem 5.1. More precisely, the optimal transmission policy can be obtained by solving the Bellman equation formulated in (5.24). To that extent, we leverage the VIA to find the optimal transmission sequence. In other words, and as it has been done before, we investigate $\Delta J'_{t+1}(S) = J'^{1}_{t+1}(S) - J'^{0}_{t+1}(S)$ where $J'^{1}_{t+1}(S)$ and $J'^{0}_{t+1}(S)$ are the value functions at time $t+1$ if $\psi(t) = 1$ and $\psi(t) = 0$ respectively. By obeying to the dynamics reported in Section 5.3.2, we get:

$$\Delta J'_{t+1}(0) = \lambda, \tag{C.9}$$

$$\Delta J'_{t+1}(S) = \lambda + p_s(p_t - p_R)(J'_t(S+1) - J'_t(0)), \quad \forall S \in \mathbb{N}^*. \tag{C.10}$$

As $\lambda \geq 0$, we can conclude that the action of remaining idle is always optimal when $S = 0$. As for the case where $S \neq 0$, we can see that $\Delta J'_{t+1}(S)$ is the sum of a positive constant and a decreasing non-positive function. Consequently, we have that the optimal action is increasing with S from $\psi^* = 0$ to $\psi^* = 1$. In other words, the difference $\Delta J'_{t+1}(S)$ decreases with S, and at a certain point, the action of transmitting becomes more beneficial than remaining idle. Therefore, we can conclude that the optimal policy of the problem is of a threshold nature.

C.4 Proof of Proposition 5.2

To proceed with the proof, we first formulate the general balance equation at state 1, which leads to $\pi_1(n) = (N-1)p_t\pi_0(n)$. Afterward, we provide the general balance equations at states k, with $2 \leq k \leq n$:

$$\pi_k(n) = \big(p_R + (N-2)p_t\big)\pi_{k-1}(n), \qquad 2 \leq k \leq n \tag{C.11}$$

By noting the results above, along with those on $\pi_1(n)$, and by carrying on with a forward induction, the results of (5.28) can be found. Next, we formulate the balance equations at states k, with $k \geq n+1$:

$$\pi_k(n) = \big(p_R p_f + (N-2)p_t + p_s p_t\big)\pi_{k-1}(n), \quad k \geq n+1 \tag{C.12}$$

By using the above results, and those of (5.28), and by proceeding with a forward induction, we can find the equations in (5.29). Lastly, we make use of the following

fundamental equality:

$$\sum_{k=0}^{+\infty} \pi_k(n) = 1. \tag{C.13}$$

By replacing $\pi_k(n)$ with their values in (C.13) and by noting the following series results:

$$\sum_{k=1}^{n} b^{k-1} = \frac{1-b^n}{1-b}, \tag{C.14}$$

$$\sum_{k=n+1}^{+\infty} a^{k-n} = \frac{a}{1-a}, \tag{C.15}$$

we can find $\pi_0(n)$, which concludes our proof.

C.5 Proof of Proposition 5.2

To calculate the average cost of the threshold policy, we first note that the cost incurred by being at state $S = k$ is nothing but the value k of the state itself. Moreover, the transmitter attempts to send a packet solely when $S \geq n$. Consequently, we have that $\overline{C}(n, \lambda) = \overline{C}(n) + \overline{C}_1(n, \lambda)$ where:

$$\overline{C}(n) = \sum_{k=1}^{+\infty} k\pi_k(n), \tag{C.16}$$

$$\overline{C}_1(n, \lambda) = \lambda \sum_{k=n}^{+\infty} \pi_k - \lambda\alpha. \tag{C.17}$$

By replacing $\pi_k(n)$ with its value from Proposition 5.2, we have that:

$$\overline{C}(n) = (N-1)p_t\pi_0\Big(\sum_{k=1}^{n} kb^{k-1} + \sum_{k=n+1}^{+\infty} b^{n-1}ka^{k-n}\Big). \tag{C.18}$$

To further simplify the above expression, we first note that the series $\sum_{k=1}^{n} kb^{k-1}$ is nothing but the derivative with respect to b of the series $\sum_{k=0}^{n} b^k = \frac{1-b^{n+1}}{1-b}$. Consequently, by deriving the expression in the right hand side, we have that:

$$\sum_{k=1}^{n} kb^{k-1} = \frac{1 + b^n(nb - n - 1)}{(1-b)^2}. \tag{C.19}$$

Next, we can address the second term of the expression in (C.18). To that extent, we proceed with a change of variables $k' = k - n$. With that being done, and by noting the fact that $\sum_{k'=1}^{+\infty} k'a^{k'} = \frac{a}{(1-a)^2}$, the expression in (5.30) can be found. By proceeding with the same series analysis, we can deduce the expression in (5.31), which concludes our proof.

C.6 Proof of Proposition 5.3

Before investigating the general scenario, we first note that the proposition is trivially true for $n = 0$. In fact, we first note that $\overline{C}(n)$ is nothing but the average penalty of a threshold policy in the unconstrained MDP case reported in Section 5.4.2. As $\overline{C}(0) = \overline{C}(1)$ (we refer the readers to the results of Theorem 5.1), we can easily verify that we have $\overline{C}(0,0) = \overline{C}(1,0)$. To tackle the case where $n \in \mathbb{N}^*$, we provide a proof that revolves around a graphical illustration of $\overline{C}(n,\lambda)$ in function of λ in Fig. C.3. To proceed in that direction, we first study in the next lemma the variation of $\overline{C}(n)$ in function of $n, \forall n \in \mathbb{N}^*$.

Lemma C.2. *The function $\overline{C}(n)$ is increasing with n.*

Proof. By considering the expression of $\overline{C}(n)$ previously reported in (5.30), we can observe that it is rather difficult to study its variations directly. To circumvent this difficulty, we recall that $\overline{C}(n)$ is nothing but the average penalty of a threshold policy in the unconstrained MDP case reported in Section 5.4.2. The dynamics of such a threshold policy is identical to the DTMC reported in Fig. 5.5. By observing the DTMC in question, we can see that the chain can only move backward due to a transition to state 0. When the transmitter does not attempt to send a packet ($S < n$), the probability of transition to state 0 is p_t. However, when the transmitter sends packets ($S \geq n$), the probability of reducing the penalty to zero is $p_f p_t + p_R p_s$. As $p_R > p_t$, we can conclude that $p_f p_t + p_R p_s > p_t$. Consequently, a packet transmission will always increase the likelihood of transitions to state 0. Based on this, we can conclude that employing a higher threshold, which leads to a smaller number of transmissions, will undoubtedly increase the average penalty. □

By using the above results, and as $\overline{C}(n,0) = \overline{C}(n)$, we can conclude that the points on the $y-$axis in Fig. C.3 move upwards as n increases. Moreover, by using the expression of $\overline{C}(n,\lambda)$ in Theorem 5.2, we can deduce that the slope of $\overline{C}(n,\lambda)$ is nothing but $A(n) - \alpha$. Since $A(n)$ decreases when the threshold n increases, we can assert that the slope of the curves $\overline{C}(n,\lambda)$ decreases with n. By combining the above two observations, we can see that for any fixed value n, the two curves $\overline{C}(n,\lambda)$ and $\overline{C}(n+1,\lambda)$ intersect at a unique point λ_0.

C.7 Proof of Theorem 5.3

To show that $n(\lambda_{n_0}) = n_0$, it is sufficient to show that for any $n \neq n_0$, we have that $\overline{C}(n,\lambda_{n_0}) \geq \overline{C}(n_0,\lambda_{n_0})$. To prove this, the first step of our analysis consists of studying the behavior of the intersection points $\lambda(n)$ as n increases. More precisely, we consider the sequence $\big(\lambda(n)\big)_{n\in\mathbb{N}}$ as the intersection point between $\overline{C}(n,\lambda)$ and $\overline{C}(n+1,\lambda)$. By using the definition in (5.36), we have that:

$$\lambda(n) = \frac{\overline{C}(n+1) - \overline{C}(n)}{A(n) - A(n+1)} \quad \forall n \in \mathbb{N} \tag{C.20}$$

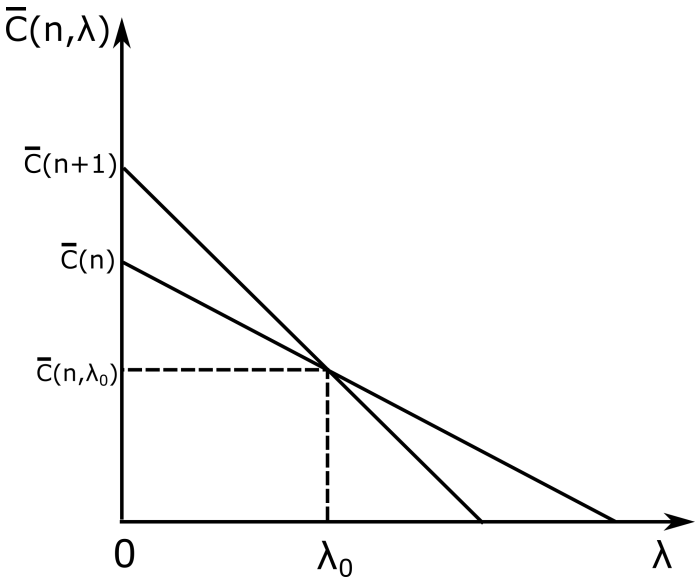


Figure C.3: Illustration of the intersection proof.

To pursue our analysis, we provide key results on the behavior of the intersection points in the following proposition.

Proposition C.1. *The sequence* $\big(\lambda(n)\big)_{n\in\mathbb{N}}$ *is increasing with n.*

Proof. As a first step, we recall that due to the results of Lemma C.2 and the decreasing nature of $A(n)$, we have that $\lambda(n) \geq 0, \ \forall n \in \mathbb{N}$. As $\overline{C}(0) = \overline{C}(1)$, we can deduce that $\lambda(0) = 0$ and, therefore, we can restrict ourselves to study the increasing property of $\big(\lambda(n)\big)_{n\in\mathbb{N}}$ solely for the case where $n \in \mathbb{N}^*$. To that extent, and as seen in Theorem 5.2, the expression of the average cost is far from trivial. Consequently, to be able to study the variations of $\big(\lambda(n)\big)_{n\in\mathbb{N}^*}$, we first provide a lemma that will be useful to our analysis.

Lemma C.3. *The series* $\big(\pi_0(n)\big)_{n\in\mathbb{N}^*}$ *is decreasing with n.*

Proof. To prove this, let us consider the series $h(n) = \frac{1}{\pi_0(n+1)} - \frac{1}{\pi_0(n)}, \ \forall n \in \mathbb{N}^*$. By replacing $\pi_0(n)$ and $\pi_0(n+1)$ by their respective values, we can show that:

$$h(n) = (N-1)p_t b^{n-1}\Big(\frac{b-a}{1-a}\Big). \tag{C.21}$$

In other words, the series $h(n)\frac{(1-a)}{(N-1)p_t(b-a)}$ is a geometric series with a common ratio b. As $a < 1$, we can conclude that the sign of $h(n)$ depends on the sign of $b-a$. To that extent, and by keeping in mind that $p_t < p_R$, we have that $b - a = p_R(1-p_f) - p_s p_t = p_s(p_R - p_t) > 0$. Hence, we can conclude that $h(n) = \frac{\pi_0(n)-\pi_0(n+1)}{\pi_0(n)\pi_0(n+1)} \geq 0, \ \forall n \in \mathbb{N}^*$. Baring in mind that $\pi_0(n) \geq 0, \ \forall n \in \mathbb{N}^*$, we can assert that $\pi_0(n) \geq \pi_0(n+1), \ \forall n \in \mathbb{N}^*$, which concludes our proof. □

With the above lemma being laid out, we now find an explicit expression of the following difference: $\Delta\overline{C} = \overline{C}(n+1) - \overline{C}(n)$. As we have previously mentioned, the

expression of the average cost is complicated, which makes treating the difference $\Delta\overline{C}$ a challenging task. To that extent, we provide in the following the 8 terms that make up $\Delta\overline{C}$:

- $z_1 = \frac{(N-1)p_t\left(\pi_0(n+1)-\pi_0(n)\right)}{(1-b)^2}$,
- $z_2 = \frac{(N-1)p_t nb^2\left(b^n\pi_0(n+1)-b^{n-1}\pi_0(n)\right)}{(1-b)^2}$,
- $z_3 = \frac{(N-1)p_t nb\left(-b^n\pi_0(n+1)+b^{n-1}\pi_0(n)\right)}{(1-b)^2}$,
- $z_4 = \frac{(N-1)p_t b^{n+1}(b-1)\pi_0(n+1)}{(1-b)^2}$,
- $z_5 = \frac{(N-1)p_t an\left(b^n\pi_0(n+1)-b^{n-1}\pi_0(n)\right)}{1-a}$,
- $z_6 = \frac{(N-1)p_t ab^n\pi_0(n+1)}{(1-a)}$,
- $z_7 = \frac{(N-1)p_t\left(-b^{n+1}\pi_0(n+1)+b^n\pi_0(n)\right)}{(1-b)^2}$,
- $z_8 = \frac{(N-1)p_t a\left(b^n\pi_0(n+1)-b^{n-1}\pi_0(n)\right)}{(1-a)^2}$.

Next, we divide each term by the expression $A(n) - A(n+1)$ previously reported in Section 5.5.4. By replacing the terms with their values, and after algebraic manipulations, we can verify that the terms that constitute the expression of $\lambda(n)$ are:

- $g_1 = \frac{z_1}{A(n)-A(n+1)} = \frac{(1-a)(-b(N-1)p_t+\frac{(N-1)p_t a(1-b)}{1-a})}{(1-b)^3(1+\frac{(N-1)p_t}{1-b})}$,
- $g(n) = \frac{\sum_{i=2}^{6} z_i}{A(n)-A(n+1)} = \frac{b-a}{1-b}(n - \frac{b}{\frac{\pi_0(n)}{\pi_0(n+1)}-b})$,
- $g_7 = \frac{z_7}{A(n)-A(n+1)} = \frac{b(1-a)}{(1-b)^2}$,
- $g_8 = \frac{z_8}{A(n)-A(n+1)} = \frac{-a}{1-a}$.

We can see that g_1, g_7, and g_8 are only constant terms. On the other hand, the term $g(n)$ requires further investigation. To that extent, we provide the following lemma.

Lemma C.4. *The series* $\left(g(n)\right)_{n\in\mathbb{N}^*}$ *is increasing with n.*

Proof. First of all, let us define the ratio $r(n)$ as $\frac{\pi_0(n)}{\pi_0(n+1)}$. To study the variations of $\big(g(n)\big)_{n\in\mathbb{N}^*}$, we consider the difference $\Delta g(n) = g(n+1) - g(n)$. By using the expression of $g(n)$, we have that:

$$\Delta g(n) = \frac{r(n)\big(r(n+1) - 2b\big) + b^2}{\big(r(n+1) - b\big)\big(r(n) - b\big)}. \tag{C.22}$$

As $r(n) \geq 1 \geq b, \ \forall n \in \mathbb{N}^*$ (we recall the results of Lemma C.3), we can conclude that it is enough to study the sign of the numerator in (C.22). By replacing $r(n)$ with its expression, we can see that to prove $\Delta g(n) \geq 0$, it is sufficient to have:

$$\frac{2b}{\pi_0(n+1)} - \frac{1}{\pi_0(n+2)} - \frac{b^2}{\pi_0(n)} \leq 0. \tag{C.23}$$

By replacing $\pi_0(n), \pi_0(n+1)$ and $\pi_0(n+2)$ with their expressions using (5.27), we can show that the Left Hand Side (LHS) of (C.23) becomes $-(b-1)^2(1 + \frac{(N-1)p_t}{1-b})$ which is always negative since $b \leq 1$. Therefore, we have that $\big(g(n)\big)_{n\in\mathbb{N}^*}$ is an increasing sequence with n. □

From the above lemma, we can conclude that $\lambda(n)$ is the sum of two terms: a constant and an increasing function with n. Therefore, the sequence $\big(\lambda(n)\big)_{n\in\mathbb{N}^*}$ is increasing with n, which concludes our proof. □

Our subsequent analysis will be divided into two sections where we study the thresholds n that are larger than n_0 and prove that they lead to a cost $\overline{C}(n, \lambda_{n_0})$ that is higher than $\overline{C}(n_0, \lambda_{n_0})$. The case where $n < n_0$ is then tackled in the section after it.

Case 1 **-** $n > n_0$: To analyze this case, we first provide the following lemma.

Lemma C.5. *$\forall k_2 > k_1$, we consider two sequences $\big(U_1(n)\big)_{n\in\mathbb{N}}$ and $\big(U_2(n)\big)_{n\in\mathbb{N}}$ such that $\big(U_2(n)\big)_{n\in\mathbb{N}^*}$ is an increasing sequence. If $\frac{U_1(n+1)-U_1(n)}{U_2(n+1)-U_2(n)}$ increases with n, then the following holds:*

$$\frac{U_1(k_2) - U_1(k_1)}{U_2(k_2) - U_2(k_1)} \geq \frac{U_1(k_1+1) - U_1(k_1)}{U_2(k_1+1) - U_2(k_1)}. \tag{C.24}$$

Proof. The proof is based on mathematical induction. More precisely, we know that the above lemma is true for $k_2 = k_1 + 1$. We suppose that it is true for any $k_2 > k_1 + 1$ and investigate the property for $k_2 + 1$. To that extent, we have that $\frac{U_1(k_2+1)-U_1(k_1)}{U_2(k_2+1)-U_2(k_1)}$ can be rewritten as:

$$\frac{U_1(k_2+1) - U_1(k_2)}{U_2(k_2+1) - U_2(k_1)} + \frac{U_1(k_2) - U_1(k_1)}{U_2(k_2+1) - U_2(k_1)}. \tag{C.25}$$

By multiplying the first and second term by $\frac{U_2(k_2+1)-U_2(k_2)}{U_2(k_2+1)-U_2(k_2)}$ and $\frac{U_2(k_2)-U_2(k_1)}{U_2(k_2)-U_2(k_1)}$ respectively, and by taking into account the increasing property of the ratio $\frac{U_1(n+1)-U_1(n)}{U_2(n+1)-U_2(n)}$ along with the induction assumption, the results can be found to be true for $k_2 + 1$, which concludes our proof. □

We can apply the above lemma by taking $U_1(n) = \overline{C}(n)$, $U_2(n) = -A(n)$ and noting the results of Proposition C.1 on $\big(\lambda(n)\big)_{n\in\mathbb{N}}$. Consequently, Lemma C.5 tells us that the intersection between $\overline{C}(n,\lambda)$ and $\overline{C}(n_0,\lambda)$ for any $n > n_0+1$ occurs after λ_{n_0}. By observing Fig. C.4, we can see that this leads to $\overline{C}(n,\lambda_{n_0})$ being larger than $\overline{C}(n_0,\lambda_{n_0})$ due to the properties of the curve $\overline{C}(n,\lambda)$ previously reported in Lemma C.2.

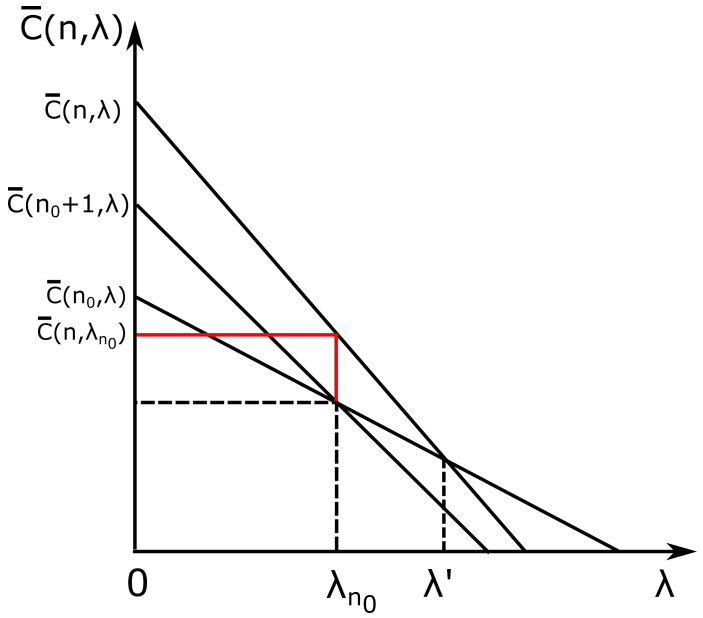


Figure C.4: Illustration of the proof: $n > n_0$.

Case 2 **-** $n < n_0$: Similarly to the previous subsection, we provide two vital lemmas to our analysis.

Lemma C.6. $\forall n \geq 1$, *if the conditions of Lemma C.5 are satisfied, we have that:*

$$\frac{U_1(n) - U_1(n-1)}{U_2(n) - U_2(n-1)} \leq \frac{U_1(n+1) - U_1(n-1)}{U_2(n+1) - U_2(n-1)}, \tag{C.26}$$

$$\frac{U_1(n+1) - U_1(n-1)}{U_2(n+1) - U_2(n-1)} \leq \frac{U_1(n+1) - U_1(n)}{U_2(n+1) - U_2(n)}. \tag{C.27}$$

Proof. We first start by rewriting $\frac{U_1(n+1)-U_1(n-1)}{U_2(n+1)-U_2(n-1)}$ as $\frac{U_1(n+1)-U_1(n)}{U_2(n+1)-U_2(n-1)} + \frac{U_1(n)-U_1(n-1)}{U_2(n+1)-U_2(n-1)}$. Afterward, the proof is based on multiplying the above expression by $\frac{U_2(n+1)-U_2(n)}{U_2(n+1)-U_2(n)}$ and $\frac{U_2(n)-U_2(n-1)}{U_2(n)-U_2(n-1)}$, and using the conditions of the lemma to prove the LHS and RHS inequalities, respectively. The details are omitted. □

Lemma C.7. $\forall n \leq n_0 - 1$, *we always have that:*

$$\frac{\overline{C}(n_0) - \overline{C}(n)}{A(n) - A(n_0)} \geq \frac{\overline{C}(n_0) - \overline{C}(n-1)}{A(n-1) - A(n_0)} \geq \frac{\overline{C}(n) - \overline{C}(n-1)}{A(n-1) - A(n)}. \tag{C.28}$$

Proof. The proof is based on a mathematical *backward* induction. As a first step, we tackle the case for $n = n_0 - 1$. As $\lambda(n)$ is increasing with n, we have that $\frac{\overline{C}(n_0)-\overline{C}(n_0-1)}{A(n_0-1)-A(n_0)} \geq \frac{\overline{C}(n_0-1)-\overline{C}(n_0-2)}{A(n_0-2)-A(n_0-1)}$. By applying Lemma C.6 for $n = n_0 - 1$, we can conclude that the above property is true for $n = n_0 - 1$. We now suppose that this property holds for any $n < n_0 - 1$ and aim to prove it to be true for $n - 1$. By using our supposition, along with the increasing property of $\lambda(n)$ and the results of Lemma C.6, the property can be verified to be true for $n - 1$, which concludes our proof. □

Equipped with the above two lemmas, we will be able to show that for any $n < n_0$, we have that $\overline{C}(n_0, \lambda_{n_0}) \leq \overline{C}(n, \lambda_{n_0})$. To do so, we aim to show that the intersection between the curves $\overline{C}(n, \lambda)$ and $\overline{C}(n_0, \lambda)$ for any $n < n_0$ occur before λ_{n_0}. Combined with the properties of the curve $\overline{C}(n, \lambda)$ previously reported in Lemma C.2, we can see in Fig. C.5 that this is equivalent to what we are aiming to prove. Our goal is, therefore, summarized in proving that: $\frac{\overline{C}(n_0+1)-\overline{C}(n_0)}{A(n_0)-A(n_0+1)} \geq \frac{\overline{C}(n_0)-\overline{C}(n)}{A(n)-A(n_0)}$ for any $n < n_0$. From the first inequality of the results of Lemma C.7, we can conclude that the series $\frac{\overline{C}(n_0)-\overline{C}(n)}{A(n)-A(n_0)}$ is increasing with n for all $n \leq n_0 - 1$. Therefore, we have that for all $n < n_0$:

$$\frac{\overline{C}(n_0) - \overline{C}(n_0 - 1)}{A(n_0 - 1) - A(n_0)} \geq \frac{\overline{C}(n_0) - \overline{C}(n)}{A(n) - A(n_0)}. \tag{C.29}$$

Lastly, by using the fact that $\lambda(n)$ is increasing with n, we can conclude that: $\frac{\overline{C}(n_0+1)-\overline{C}(n_0)}{A(n_0)-A(n_0+1)} \geq \frac{\overline{C}(n_0)-\overline{C}(n_0-1)}{A(n_0-1)-A(n_0)}$. Combining this with the results of eq. (C.29), we can conclude our proof.

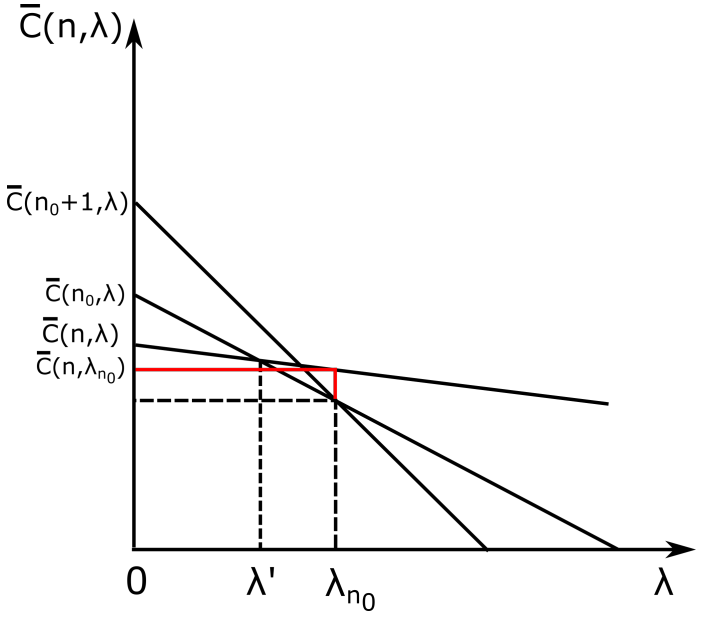


Figure C.5: Illustration of the proof: $n < n_0$.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre: Optimisation des Réseaux Sans-Fil: Fraîcheur en Communication

Mots clés: Âge de l'information, Paquets de Mise à Jour, Estimation à Distance, Âge de l'Information Incorrecte, Accès Multiple Avec Écoute de la Porteuse, CSMA, Planification Distribuée, Mise en File d'Attente Prioritaire

Résumé: La prolifération des smartphones, avec la connectivité omniprésente et le bas coût des matériaux, a ouvert la voie à de nouvelles applications qui reposent sur la livraison en temps opportun des paquets d'un bout à l'autre du réseau. De la surveillance des appareils ménagers à la maison au réseaux de véhicules où les informations de vitesse et de position du véhicule sont diffusées, ces applications nécessitent de données frais pour avoir des performances optimales. Pour quantifier cette notion de fraîcheur de données, le concept de l'Âge de l'Information (AdI) est né, et la recherche s'est fortement concentrée sur son analyse et son optimisation dans divers contextes de réseau. Cette thèse explore l'AdI dans de nombreux environnements, met en lumière ses points faibles et leur apporte ainsi des solutions dans plusieurs applications de surveillance en temps réel.

Dans la première partie de la thèse, nous nous concentrons sur l'optimisation des métriques basées sur l'âge dans les systèmes de communication fondamentaux. Plus précisément, dans le troisième chapitre, nous examinons les métriques basées sur l'âge dans les environnements multi-classes qui sont abondants dans les applications en temps réel. Un exemple simple est celui des réseaux de véhicules où les données relatives à la sécurité sont considérées comme plus sensibles. Par conséquent, elles ont une priorité plus élevée par rapport aux autres données du système. Nous dérivons une expression de l'âge moyen de chaque flux et nous fournissons des résultats à propos de l'interaction entre les multiples classes. Cela ouvre la voie à la deuxième partie du chapitre, où nous introduisons un nouveau cadre d'optimisation basé sur l'AdI dans les systèmes multi-classes. Nous y caractérisons les gains en termes de fraîcheur de l'information lorsque notre cadre est adopté par rapport à des approches de pointe. Le quatrième chapitre traite un environnement distribué, où les appareils accèdent au canal en utilisant la méthode d'accès multiple avec écoute de la porteuse (CSMA). CSMA est considéré comme l'un des méthodes d'accès canal distribués les plus connus et les plus répandus (par exemple, CSMA est le principal moyen d'accès en Wi-Fi). Dans ce cas, nous caractérisons, grâce à des analyses théoriques rigoureuses, le point de fonctionnement optimal qui minimise l'âge moyen du réseau.

Dans la deuxième partie de la thèse, nous mettons en lumière les lacunes de l'âge de l'information et des métriques d'erreur standard dans de nombreuses applications en temps réel. Par conséquent, nous introduisons une nouvelle métrique de performance, que nous appelons l'Âge de l'Information Incorrecte (AdII). L'AdII traite ces lacunes en étendant la notion de données frais et en saisissant l'effet de détérioration que les informations incorrectes peuvent avoir avec le temps sur le système. Dans les scénarios à la fois sans et avec contraintes de ressources, nous dérivons des politiques d'échantillonnage optimales qui minimisent l'AdII. Nous soulignons également leurs avantages par rapport aux politiques optimales pour l'âge et pour les métriques d'erreur standard dans diverses applications. Nos résultats et analyses fournissent des informations clés sur la métrique d'âge et ouvrent la voie à de nouvelles orientations de recherche pour les applications de surveillance en temps réel.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Title: Optimization of Wireless Networks: Freshness in Communications

Keywords: Age of Information, Status Updates, Remote Estimation, Age of Incorrect Information, Carrier Sense Multiple Access, CSMA, Distributed Scheduling, Priority Queuing

Abstract: The proliferation of smartphones, along with the ubiquitous connectivity and cheap hardware cost, has paved the way for new applications that rely on the timely delivery of packets from one end of the network to another. From monitoring home appliances back at the house to vehicular networks where the vehicle's velocity and position information are disseminated, these applications require fresh data to have optimal performance. To quantify this notion of freshness, the concept of the Age of Information (AoI) was born, and research attention has been put heavily on its analysis and optimization in various network settings. This thesis explores the AoI in numerous system environments, sheds light on its shortcomings, and accordingly provides solutions to them in several real-time monitoring applications.

In the first part of the thesis, we focus on optimizing age-based metrics in fundamental communication systems. Specifically, in the third chapter, we examine age-based metrics in multi-class environments that are abundant in real-time applications. A simple example is vehicular networks where safety-related data are considered more sensitive. Consequently, they have a higher priority than the other data in the system. We derive a closed-form expression of each stream's average age and provide substantial insights into the interaction between the multiple classes. This paves the way for the second part of the chapter, where we introduce a new AoI-based optimization framework in multi-class systems. Therein, we characterize the gains in terms of information freshness when our framework is adopted compared to state-of-the-art approaches. The fourth chapter deals with a distributed scheduling environment, where devices contend for the channel using the well-known carrier sense multiple access scheme (CSMA). CSMA is considered one of the most renowned and widely spread distributed scheduling schemes (e.g., CSMA is the primary medium access in Wi-Fi). We characterize, through rigorous theoretical analyses, the operating point that minimizes the average AoI.

In the second part of the thesis, we shed light on the shortcomings of the age of information and standard error metrics in many real-time applications. Toward that end, we introduce a new performance metric, which we refer to as the Age of Incorrect Information (AoII). AoII deals with these shortcomings as it extends the notion of fresh updates and captures the deteriorating effect wrong information can have with time on the system. In both unconstrained and resource-constraint scenarios, we derive optimal sampling policies that minimize the AoII. We also highlight their advantages compared to both the age-optimal and error-optimal policies in a variety of real-life applications. Our results and analyses provide key insights into the age metric and lead the way to novel research directions for real-time monitoring applications.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France