entropy

MDPI

Article

Scheduling to Minimize Age of Incorrect Information with Imperfect Channel State Information

Yutao Chen and Anthony Ephremides *

Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA; cheny@umd.edu

* Correspondence: etony@umd.edu

Abstract: In this paper, we study a slotted-time system where a base station needs to update multiple users at the same time. Due to the limited resources, only part of the users can be updated in each time slot. We consider the problem of minimizing the Age of Incorrect Information (AoII) when imperfect Channel State Information (CSI) is available. Leveraging the notion of the Markov Decision Process (MDP), we obtain the structural properties of the optimal policy. By introducing a relaxed version of the original problem, we develop the Whittle's index policy under a simple condition. However, indexability is required to ensure the existence of Whittle's index. To avoid indexability, we develop Indexed priority policy based on the optimal policy for the relaxed problem. Finally, numerical results are laid out to showcase the application of the derived structural properties and highlight the performance of the developed scheduling policies.

Keywords: age of incorrect information; multi-user system; scheduling policy

Citation: Chen, Y.; Ephremides, A. Scheduling to Minimize Age of Incorrect Information with Imperfect Channel State Information. *Entropy* **2021**, *23*, 1572. https://doi.org/10.3390/e23121572

Academic Editor: Mario Martinelli

Received: 3 November 2021
Accepted: 23 November 2021
Published: 25 November 2021

1. Introduction

The Age of Incorrect Information (AoII) is introduced in [1] as a combination of age-based metrics (e.g., Age of Information (AoI)) and error-based metrics (e.g., Minimum Mean Square Error). In communication systems, AoII captures not only the information mismatch between the source and the destination but also the aging process of inconsistent information. Hence, two functions dominate AoII. The first is the time penalty function, which reflects how the inconsistency of information affects the system over time. In real-life applications, inconsistent information will affect different communication systems in different ways. For example, machine temperature monitoring is time-sensitive because the damage caused by overheating will accumulate quickly. However, reservoir water level monitoring is less sensitive to time. Therefore, by adopting different time penalty functions, AoII can capture different aging processes of the mismatch in different systems. The second is the information penalty function, which captures the information mismatch between the source and the destination. It allows us to measure mismatches in different ways, depending on how sensitive different systems are to information inconsistencies. For example, the navigation system requires precise information to give correct instructions, but the real-time delivery tracking system does not need very accurate location information. Since we can choose different penalty functions for different systems, AoII is adaptable to various communication goals, which is why it is regarded as a semantic metric [2].

Since the introduction of AoII, several studies have been performed to reveal its fundamental nature. The authors of [3] consider a system with random packet delivery times and compare AoII with AoI and real-time error via extensive numerical results. The authors of [4] study the problem of minimizing the AoII that takes the general time penalty function. Three real-life applications are considered to showcase the performance advantages of AoII over AoI and real-time error. In [5], the authors investigate the AoII that considers the quantified mismatch between the source and the destination. The optimization problem is studied when the system is resource-constrained. The authors of [6] studied the AoII

minimization problem in the context of scheduling. It considers a system where the central scheduler needs to update multiple users at the same time. However, the central scheduler cannot know the states of the sources before receiving the updates. By introducing the belief value, Whittle's index policy is developed and evaluated. In this paper, we also consider the problem of minimizing AoII in scheduling. Different from [6], we consider the generic time penalty function and study the minimization problem in the presence of imperfect Channel State Information (CSI). Due to the existence of CSI, Whittle's index policy becomes infeasible in general. Hence, we introduce another scheduling policy that is more versatile and has comparable performance to Whittle's index policy.

The problem of scheduling to minimize AoI is studied under various system settings in [7–11]. The problem studied in this paper is different and more complicated because AoII considers the aging process of inconsistent information rather than the aging process of updates. Meanwhile, none of them consider the case where CSI is available. The problem of optimizing information freshness in the presence of CSI is studied in [12,13]. However, they focus on the system with a single user and mainly discuss the case where CSI is perfect. The scheduling problems with the goal of minimizing an error-based performance measure are considered in [14–16]. Our problem is fundamentally different because AoII also considers the time effect. Moreover, we consider the system where a base station observes multiple sources simultaneously and needs to send updates to multiple destinations.

The main contributions of this work can be summarized as follows. (1) We study the problem of minimizing AoII in a multi-user system where imperfect CSI is available. Meanwhile, the time penalty function is generic. (2) We derive the structural properties of the optimal policy for the considered problem. (3) We establish the indexability of the considered problem under a simple condition and develop Whittle's index policy. (4) We obtain the optimal policy for a relaxed version of the original problem. By exploring the characteristics of the relaxed problem, we provide an efficient algorithm to obtain the optimal policy. (5) Based on the optimal policy for the relaxed problem, we develop the Indexed priority policy that is free from indexability and has comparable performance to Whittle's index policy.

The remainder of this paper is organized in the following way. In Section 2, we introduce the system model and formulate the primal problem. Section 3 explores the structural properties of the optimal policy for the primal problem. Under a simple condition, we develop Whittle's index policy in Section 4. Section 5 presents the optimal policy for a relaxed version of the primal problem. On this basis, we develop the Indexed priority policy in Section 6. Finally, in Section 7, the numerical results are laid out.

2. System Overview

2.1. Communication Model

We consider a slotted-time system with N users and one base station. Each user is composed of a source process, a channel, and a receiver. We assume all the users share the same structure, but the parameters are different. The structure of the communication model is provided in Figure 1.

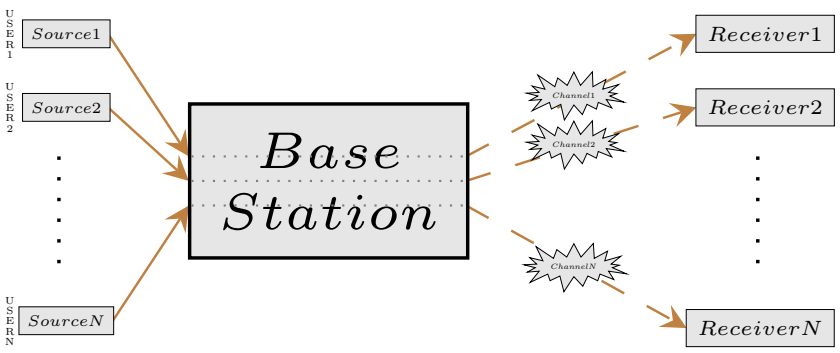


Figure 1. The structure of the communication model.

For user i, the source process is modeled by a two-state Markov chain where transitions happen between the two states with probability $p_i > 0$ and self-transitions happen with probability $1 - p_i$. At any time slot t, the state of the source process $X_{i,t} \in \{0, 1\}$ will be reported to the base station as an update, and the base station will decide whether to transmit this update through the corresponding channel. The channel is unreliable, but the estimate of the Channel State Information (CSI) is available at the beginning of each time slot. Let $r_{i,t} \in \{0, 1\}$ be the CSI at time t. We assume that $r_{i,t}$ is independent across time and user indices. $r_{i,t} = 1$ if and only if the transmission attempt at time t will succeed and $r_{i,t} = 0$ otherwise. Then, we denote by $\hat{r}_{i,t} \in \{0, 1\}$ the estimate of $r_{i,t}$. We assume that $\hat{r}_{i,t}$ is an independent Bernoulli random variable with parameter γ_i, i.e., $\hat{r}_{i,t} = 1$ with probability $\gamma_i \in [0, 1]$ and $\hat{r}_{i,t} = 0$ with probability $1 - \gamma_i$. However, the estimate is imperfect. We assume that the error depends only on the user and its estimate. More precisely, we define the probability of error as $p_{e,i}^{\hat{r}_i} \triangleq Pr[r_i \neq \hat{r}_i \mid \hat{r}_i]$. We assume $p_{e,i}^{\hat{r}_i} < 0.5$ because we can flip the estimate if $p_{e,i}^{\hat{r}_i} > 0.5$. We are not interested in the case of $p_{e,i}^{\hat{r}_i} = 0.5$ since $\hat{r}_{i,t}$ is useless in this case. Although the channel is unreliable, each transmission attempt takes exactly one time slot regardless of the result, and the successfully transmitted update will not be corrupted. Every time an update is received, the receiver will use it as the new estimate $\hat{X}_{i,t}$. The receiver will send an $ACK/NACK$ packet to inform the base station of its reception of the new update. Since an $ACK/NACK$ packet is generally very small and simple, we assume that it is transmitted reliably and received instantaneously. Then, if ACK is received, the base station knows that the receiver's estimate changed to the transmitted update. If $NACK$ is received, the base station knows that the receiver's estimate did not change. Therefore, the base station always knows the estimate at the receiver side.

At the beginning of each time slot, the base station receives updates from each source and the estimates of CSI from each channel. The old updates and estimates are discarded upon the arrival of new ones. Then, the base station decides which updates to transmit, and the decision is independent of the transmission history. Due to the limited resources, at most $M < N$ updates are allowed per transmission attempt. We consider a base station that always transmits M updates.

2.2. Age of Incorrect Information

All the users adopt AoII as a performance metric, but the choices of penalty functions vary. Let X_t and $\hat{X}_t$ be the true state and the estimate of the source process, respectively. Then, in a slotted-time system, AoII can be expressed as follows

$$\Delta_{AoII}(X_t, \hat{X}_t, t) = \sum_{k=U_t+1}^{t} \big(g(X_k, \hat{X}_k) \times F(k - U_t)\big), \tag{1}$$

where U_t is the last time instance before time t (including t) that the receiver's estimate is correct. $g(X_t, \hat{X}_t)$ can be any information penalty function that captures the difference between X_t and $\hat{X}_t$. $F(t) \triangleq f(t) - f(t-1)$ where $f(t)$ can be any time penalty function that is non-decreasing in t. We consider the case where the users adopt the same information penalty function $g(X_t, \hat{X}_t) = |X_t - \hat{X}_t|$ but possibly different time penalty functions. To ease the analysis, we require $f(t)$ to be unbounded. Combined together, we require $f(t_1) \leq f(t_2)$ if $t_1 < t_2$ and $\lim_{t \to +\infty} f(t) = +\infty$. Without a loss of generality, we assume $f(0) = 0$, as the source is modeled by a two-state Markov chain, $g(X_t, \hat{X}_t) \in \{0, 1\}$. Hence, Equation (1) can be simplified to

$$\Delta_{AoII}(X_t, \hat{X}_t, t) = \sum_{k=U_t+1}^{t} F(k - U_t) = f(s_t),$$

where $s_t \triangleq t - U_t$. Therefore, the evolution of s_t is sufficient to characterize the evolution of AoII. To this end, we distinguish between the following cases.

- When the receiver's estimate is correct at time $t+1$, we have $U_{t+1} = t+1$. Then, by definition, $s_{t+1} = 0$.
- When the receiver's estimate is incorrect at time $t+1$, we have $U_{t+1} = U_t$. Then, by definition, $s_{t+1} = t+1-U_t = s_t+1$.

To sum up, we get

$$s_{t+1} = \mathbb{1}_{\{U_{t+1} \neq t+1\}} \times (s_t + 1). \tag{2}$$

A sample path of s_t is shown in Figure 2. In the remainder of this paper, we use $f_i(\cdot)$ to denote the time penalty function user i adopts.

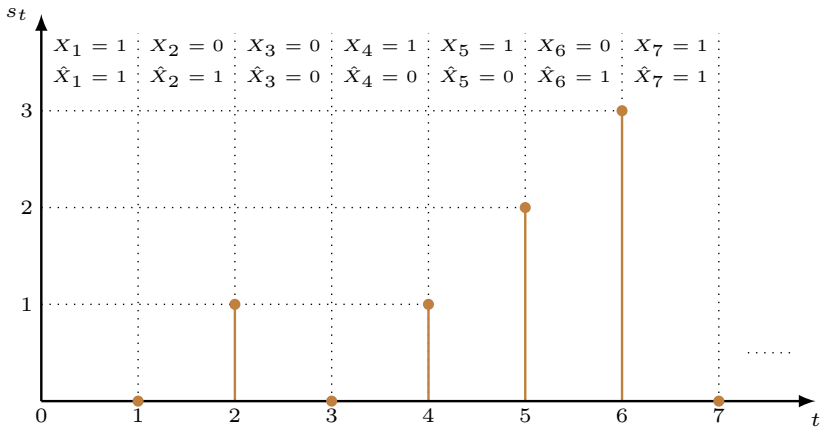


Figure 2. A sample path of s_t.

Remark 1. *Under this particular choice of the penalty function, s_t can be interpreted as the time elapsed since the last time the receiver's estimate is correct. Please note that s_t is different from the Age of Information (AoI) [17], which is defined as the time elapsed since the generation time of the last received update. We can see that AoI considers the aging process of the update, while AoII considers the aging process of the estimation error. At the same time, s_t is also fundamentally different from the holding time, which, according to [18,19], is defined as the time elapsed since the last successful transmission. We notice that the receiver's estimate can become correct even when no new update is successfully transmitted. Moreover, the information carried by the update may have become incorrect by the time it is received. We also notice that [18,19] consider the problem of minimizing the estimation error. However, by adopting AoII as the performance metric, we study the impact of estimation error on the system.*

2.3. System Dynamic

In this section, we tackle the system dynamic. We notice that the status of user i can be captured by the pair $x_{i,t} \triangleq (s_{i,t}, \hat{r}_{i,t})$. In the following, we will use $x_{i,t}$ and $(s_{i,t}, \hat{r}_{i,t})$ interchangeably. Then, the system dynamic can be fully characterized by the dynamic of $\boldsymbol{x}_t \triangleq (x_{1,t}, \ldots, x_{N,t})$. Hence, it suffices to characterize the value of $\boldsymbol{x}_{t+1}$ given $\boldsymbol{x}_t$ and the base station's action. To this end, we denote, by $\boldsymbol{a}_t = (a_{1,t}, \ldots, a_{N,t})$, the base station's action at time t. $a_{i,t} = 1$ if the base station transmits the update from user i at time t and $a_{i,t} = 0$ otherwise. We notice that given action $\boldsymbol{a}_t$, users are independent and the action taken on user i will only affect itself. Consequently

$$Pr(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t, \boldsymbol{a}_t) = \prod_{i=1}^{N} Pr(x_{i,t+1} \mid x_{i,t}, \boldsymbol{a}_t) = \prod_{i=1}^{N} Pr(x_{i,t+1} \mid x_{i,t}, a_{i,t}).$$

Combined with the fact that all the users share the same structure, it is sufficient to study the dynamic of a single user. In the following discussions, we drop the user-dependent subscript i. We recall that $\hat{r}_{t+1}$ is an independent Bernoulli random variable. Then, we have

$$Pr(x_{t+1} \mid x_t, a_t) = P(\hat{r}_{t+1}) \times Pr(s_{t+1} \mid x_t, a_t). \tag{3}$$

By definition, $P(\hat{r}_{t+1} = 1) = \gamma$ and $P(\hat{r}_{t+1} = 0) = 1 - \gamma$. Then, we only need to tackle the value of $Pr(s_{t+1} \mid x_t, a_t)$. To this end, we distinguish between the following cases

- When $x_t = (0, \hat{r}_t)$, the estimate at time t is correct (i.e., $\hat{X}_t = X_t$). Hence, for the receiver, X_t carries no new information about the source process. In other words, $\hat{X}_{t+1} = \hat{X}_t$ regardless of whether an update is transmitted at time t. We recall that $U_{t+1} = U_t$ if $\hat{X}_{t+1} \neq X_{t+1}$ and $U_{t+1} = t + 1$ otherwise. Since the source is binary, we obtain $U_{t+1} = U_t$ if $X_{t+1} \neq X_t$, which happens with probability p and $U_{t+1} = t + 1$ otherwise. According to (2), we obtain

$$Pr(1 \mid (0, \hat{r}_t), a_t) = p,$$
$$Pr(0 \mid (0, \hat{r}_t), a_t) = 1 - p.$$

- When $a_t = 0$ and $x_t = (s_t, \hat{r}_t)$, where $s_t > 0$, the channel will not be used and no new update will be received by the receiver,and so, $\hat{X}_{t+1} = \hat{X}_t$. We recall that $U_{t+1} = U_t$ if $\hat{X}_{t+1} \neq X_{t+1}$ and $U_{t+1} = t + 1$ otherwise. Since $X_t \neq \hat{X}_t$ and the source is binary, we have $U_{t+1} = U_t$ if $X_{t+1} = X_t$, which happens with probability $1 - p$ and $U_{t+1} = t + 1$ otherwise. According to (2), we obtain

$$Pr(s_t + 1 \mid (s_t, \hat{r}_t), a_t = 0) = 1 - p,$$
$$Pr(0 \mid (s_t, \hat{r}_t), a_t = 0) = p.$$

- When $a_t = 1$ and $x_t = (s_t, 1)$ where $s_t > 0$, the transmission attempt will succeed with probability $1 - p_e^1$ and fail with probability p_e^1. We recall that $U_{t+1} = U_t$ if $\hat{X}_{t+1} \neq X_{t+1}$ and $U_{t+1} = t + 1$ otherwise. Then, when the transmission attempt succeeds (i.e., $\hat{X}_{t+1} = X_t$), $U_{t+1} = U_t$ if $X_{t+1} \neq X_t$ and $U_{t+1} = t + 1$ otherwise. When the transmission attempt fails (i.e., $\hat{X}_{t+1} = \hat{X}_t \neq X_t$), we have $U_{t+1} = U_t$ if $X_{t+1} = X_t$ and $U_{t+1} = t + 1$ otherwise. Combining (2) with the dynamic of the source process we obtain

$$Pr(s_t + 1 \mid (s_t, 1), a_t = 1) = p_e^1(1 - p) + (1 - p_e^1)p \triangleq \alpha,$$
$$Pr(0 \mid (s_t, 1), a_t = 1) = p_e^1 p + (1 - p_e^1)(1 - p) = 1 - \alpha.$$

- When $a_t = 1$ and $x_t = (s_t, 0)$, where $s_t > 0$, following the same line, we obtain

$$Pr(s_t + 1 \mid (s_t, 0), a_t = 1) = p_e^0 p + (1 - p_e^0)(1 - p) \triangleq \beta,$$
$$Pr(0 \mid (s_t, 0), a_t = 1) = p_e^0(1 - p) + (1 - p_e^0)p = 1 - \beta.$$

Combines together, we obtain the value of $Pr(s_{t+1} \mid x_t, a_t)$ in all cases. As only M out of N updates are allowed per transmission attempt, we realize a necessity to require transmission attempts always help minimize AoII. It is equivalent to impose $Pr(s_{t+1} > s_t \mid (s_t, \hat{r}_t), a_t = 0) > Pr(s_{t+1} > s_t \mid (s_t, \hat{r}_t), a_t = 1)$ for any $(s_t, \hat{r}_t)$. Leveraging the results above, it is sufficient to require $p < 0.5$. As all the users share the same structure, we assume, for the rest of this paper, that $0 < p_i < 0.5$ for $1 \leq i \leq N$.

2.4. Problem Formulation

The communication goal is to minimize the expected AoII. Therefore, the problem can be formulated as the following

$$\underset{\phi \in \Phi}{\arg\min} \quad \lim_{T \to \infty} \frac{1}{T} \mathbb{E}_\phi \left(\sum_{t=0}^{T-1} \sum_{i=1}^{N} f_i(s_{i,t}) \right) \tag{4a}$$

$$\text{subject to} \quad \sum_{i=1}^{N} a_{i,t} = M \quad \forall t, \tag{4b}$$

where Φ is the set of all causal policies. We refer to the constrained minimization problem reported in problem (4) as the Primal Problem (PP). We notice that the PP is a Restless Multi-Armed Bandit (RMAB) Problem. The optimal policy for this type of problem is far from reachable since it is PSPACE-hard in general [20]. However, we can still derive the structural properties of the optimal policy. These structural properties can be used as a guide for the development of scheduling policies and can indicate the good performance of the developed scheduling policies.

3. Structural Properties of the Optimal Policy

In this section, we investigate the structural properties of the optimal policy for PP. We first define an infinite horizon with an average cost Markov Decision Process (MDP) $\mathcal{M}_N(w, M) = (\mathcal{X}_N, \mathcal{A}_N(M), \mathcal{P}_N, \mathcal{C}_N(w))$, where

- $\mathcal{X}_N$ denotes the state space. The state is $\boldsymbol{x} = (x_1, \ldots, x_N)$ where $x_i = (s_i, \hat{r}_i)$.
- $\mathcal{A}_N(M)$ denotes the action space. The feasible action is $\boldsymbol{a} = (a_1, \ldots, a_N)$ where $a_i \in \{0, 1\}$ and $\sum_{i=1}^{N} a_i = M$. Note that the feasible actions are independent of the state and the time.
- $\mathcal{P}_N$ denotes the state transition probabilities. We define $P_{\boldsymbol{x},\boldsymbol{x}'}(\boldsymbol{a})$ as the probability that action $\boldsymbol{a}$ at state $\boldsymbol{x}$ will lead to state $\boldsymbol{x}'$. It is calculated by

$$P_{\boldsymbol{x},\boldsymbol{x}'}(\boldsymbol{a}) = \prod_{i=1}^{N} P(\hat{r}_i') P_{s_i,s_i'}(a_i, \hat{r}_i),$$

  where $P_{s_i,s_i'}(a_i, \hat{r}_i)$ is the transition probability from s_i to s_i' when the estimate of CSI is $\hat{r}_i$ and action a_i is taken. The values of $P_{s_i,s_i'}(a_i, \hat{r}_i)$ can be obtained easily from the results in Section 2.3.
- $\mathcal{C}_N(w)$ denotes the instant cost. When the system is at state $\boldsymbol{x}$ and action $\boldsymbol{a}$ is taken, the instant cost is $C(\boldsymbol{x}, \boldsymbol{a}) \triangleq \sum_{i=1}^{N} C(x_i, a_i) \triangleq \sum_{i=1}^{N} \left(f_i(s_i) + w a_i\right)$.

We notice that PP can be cast into $\mathcal{M}_N(0, M)$. Since $w = 0$, the instant cost is independent of action $\boldsymbol{a}$. Therefore, we abbreviate $C(\boldsymbol{x}, \boldsymbol{a})$ as $C(\boldsymbol{x})$. To simplify the analysis, we consider the case of $M = 1$. Equivalently, we investigate the structural properties of the optimal policy for $\mathcal{M}_N(0, 1)$.

Remark 2. *For the case of $M > 1$, we can apply the same methodology. However, as M increases, the action space will grow quickly, resulting in the need to consider more feasible actions in each step of the proof. Hence, to better demonstrate the methodology, we only consider the case of $M = 1$ in this paper.*

It is well known that the optimal policy for $\mathcal{M}_N(0, 1)$ can be characterized by the value function. We denote the value function of state $\boldsymbol{x}$ as $V(\boldsymbol{x})$. A canonical procedure to calculate $V(\boldsymbol{x})$ is applying the Value Iteration Algorithm (VIA). To this end, we define $V_\nu(\cdot)$ as the estimated value function at iteration ν of VIA and initialize $V_0(\cdot) = 0$. Then, VIA updates the estimated value functions in the following way

$$V_{\nu+1}(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \min_{\boldsymbol{a} \in \mathcal{A}_N(1)} \left\{ \sum_{\boldsymbol{x}' \in \mathcal{X}_N} P_{\boldsymbol{x},\boldsymbol{x}'}(\boldsymbol{a}) V_\nu(\boldsymbol{x}') \right\}, \tag{5}$$

where θ is the optimal value of $\mathcal{M}_N(0, 1)$. VIA is guaranteed to converge to the value function [21]. More precisely, $V_\nu(\cdot) = V(\cdot)$ when $\nu \to +\infty$. However, the exact value function is impossible to get since we need infinite iterations and the state space is infinite. Instead, we provide two structural properties of the value function.

Lemma 1 (Monotonicity)**.** *For $\mathcal{M}_N(0, 1)$, $V(\boldsymbol{x})$ is non-decreasing in s_i for $1 \leq i \leq N$.*

Proof. Leveraging the iterative nature of VIA, we use mathematical induction to prove the desired results. The complete proof can be found in Appendix A. □

Before introducing the next structural property, we make the following definition.

Definition 1 (Statistically identical)**.** *Two users are said to be statistically identical if the user-dependent parameters and the adopted time penalty functions are the same.*

For the users that are statistically identical, we can prove the following

Lemma 2 (Equivalence)**.** *For $\mathcal{M}_N(0,1)$, if users j and k are statistically identical, $V(\boldsymbol{x}) = V(\mathscr{P}(\boldsymbol{x}))$ where $\mathscr{P}(\boldsymbol{x})$ is state $\boldsymbol{x}$ with x_j and x_k exchanged.*

Proof. Leveraging the iterative nature of VIA, we use mathematical induction to prove the desired results. At each iteration, we show that for each feasible action at state $\boldsymbol{x}$, we can find an equivalent action at state $\mathscr{P}(\boldsymbol{x})$. Two actions are equivalent if they lead to the same value function. The complete proof can be found in Appendix B. □

Equipped with the above lemmas, we proceed with characterizing the structural properties of the optimal policy. We recall that the optimal action at each state can be characterized by the value function. Hence, we denote, by $V^j(\boldsymbol{x})$, the value function resulting from choosing user j to update at state $\boldsymbol{x}$. Then, $V^j(\boldsymbol{x})$ can be calculated by

$$V^j(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}'-x'_j} \left\{ \left(\prod_{i \neq j} P_{x_i,x'_i}(0) \right) \sum_{\hat{r}'_j} \left[P(\hat{r}'_j) \left(\sum_{s'_j} P_{s_j,s'_j}(1,\hat{r}_j) V(\boldsymbol{x}') \right) \right] \right\}.$$

If $V^j(\boldsymbol{x}) < V^k(\boldsymbol{x})$ for all $k \neq j$, it is optimal to transmit the update from user j. When $V^j(\boldsymbol{x}) = V^k(\boldsymbol{x})$, the two choices are equally desirable. In the following, we will characterize the properties of $\delta^{j,k}(\boldsymbol{x}) \triangleq V^j(\boldsymbol{x}) - V^k(\boldsymbol{x})$ for any j and k.

Theorem 1 (Structural properties)**.** *For $\mathcal{M}_N(0,1)$, $\delta^{j,k}(\boldsymbol{x})$ has the following properties*

1. *$\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $\hat{r}_k = p^0_{e,k} = 0$. The equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$.*
2. *$\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $\hat{r}_j$ and is non-decreasing in $\hat{r}_k$ when $s_j, s_k > 0$. At the same time, $\delta^{j,k}(\boldsymbol{x})$ is independent of $\hat{r}_i$ for any $i \neq j,k$.*
3. *$\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $s_k = 0$. The equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$.*
4. *$\delta^{j,k}(\boldsymbol{x})$ is non-increasing in s_j if $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ and is non-decreasing in s_k if $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$ when $s_j, s_k > 0$. We define $\Gamma_i^1 \triangleq \frac{\alpha_i}{1-p_i}$ and $\Gamma_i^0 \triangleq \frac{\beta_i}{1-p_i}$ for $1 \leq i \leq N$.*
5. *$\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $s_j \geq s_k$, $\hat{r}_j \geq \hat{r}_k$, and users j and k are statistically identical.*

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

We notice that $\Gamma_i^{\hat{r}_i}$ can be written as

$$\Gamma_i^{\hat{r}_i} = \frac{Pr(s_i + 1 \mid (s_i, \hat{r}_i), a_i = 1)}{Pr(s_i + 1 \mid (s_i, \hat{r}_i), a_i = 0)} < 1,$$

where s_i can be any positive integer. Consequently, $\Gamma_i^{\hat{r}_i}$ is independent of any $s_i > 0$ and indicates the decrease in the probability of increasing s_i caused by action $a_i = 1$. When $\Gamma_i^{\hat{r}_i}$ is large, action $a_i = 1$ will achieve a small decrease in the probability of increasing s_i. In the following, we provide an intuitive interpretation of why the monotonicity in Property 4 of Theorem 1 depends on $\Gamma_i^{\hat{r}_i}$. We take the case of $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ as an example and assume that there are only users j and k in the system. Then, according to Section 2.3, the dynamic of s_j and s_k can be divided into the following three cases

- Neither s_j nor s_k increases. In this case, both s_j and s_k become zero.
- Either s_j or s_k increases and the other becomes zero. We denote by P_j^k the probability that only s_k increases when $a_j = 1$. The notation for other cases is defined analogously. The probabilities can be obtained easily using the results in Section 2.3.
- Both s_j and s_k increase. We denote by P_j the probability that both s_j and s_k increase when $a_j = 1$. P_k is defined analogously. The probabilities can be obtained easily using the results in Section 2.3.

We notice that $\delta^{j,k}(\boldsymbol{x})$ implies the tendency of the base station to choose between the two users. The larger $\delta^{j,k}(\boldsymbol{x})$ is, the more the base station tends to choose user k. Thus, we investigate the base station's propensity to choose user k when s_k increases but s_j stays the same. We ignore the case where the resulting s_k is zero since it is independent of the increase in s_k. With this in mind, we first notice that $P_k^k \leq P_j^k$. Meanwhile, we can easily verify that $\frac{P_j}{P_k} = \frac{\Gamma_j^{\hat{r}_j}}{\Gamma_k^{\hat{r}_k}}$. When $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$, we have $P_j \leq P_k$. Then, there exists a subtle trade-off. More precisely, choosing user k will result in $P_k^k \leq P_j^k$, but at the cost of $P_k \geq P_j$. Hence, in this case, the propensity of the base station is hard to determine. Following the same line, we can show that choosing user j will lead to $P_j^j \leq P_k^j$ and $P_j \leq P_k$. Thus, there exists no such trade-off when we investigate the base station's propensity to choose user j as s_j increases but s_k stays the same.

Leveraging Theorem 1, we can provide some specific structural properties of the optimal policy.

Corollary 1 (Application of Theorem 1)**.** *When $M = 1$, the optimal policy for PP must satisfy the following*

1. *The user i with $\hat{r}_i = p_{e,i}^0 = 0$ or $s_i = 0$ will not be chosen unless it is to break the tie.*
2. *When user j is chosen at state $\boldsymbol{x}_1$, then for state $\boldsymbol{x}_2$, such that $\hat{r}_{1,j} \leq \hat{r}_{2,j}$ and $s_{1,i} = s_{2,i}$ for $1 \leq i \leq N$, the optimal choice must be in the set $G = \{j\} \cup \{k : \hat{r}_{1,k} < \hat{r}_{2,k}\}$.*
3. *When $N = 2$, we consider two states, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, which differ only in the value of s_j. Specifically, $s_{1,j} \leq s_{2,j}$. If user j is chosen at state $\boldsymbol{x}_1$ and $\Gamma_j^{\hat{r}_{1,j}} \leq \Gamma_k^{\hat{r}_{1,k}}$, the optimal choice at state $\boldsymbol{x}_2$ will also be user j.*
4. *When $N = 2$, we consider two states, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, which differ only in the value of s_k. Specifically, $s_{1,k} \geq s_{2,k}$. If user j is chosen at state $\boldsymbol{x}_1$ and $\Gamma_j^{\hat{r}_{1,j}} \geq \Gamma_k^{\hat{r}_{1,k}}$, the optimal choice at state $\boldsymbol{x}_2$ will also be user j.*
5. *When all users are statistically identical, the optimal choice at any time slot must be either the user with $x = (s_{max,1}, 1)$ where $s_{max,1} \triangleq \max_{s_i}\{(s_i, 1)\}$ or the user with $x = (s_{max,0}, 0)$ where $s_{max,0} \triangleq \max_{s_i}\{(s_i, 0)\}$. Moreover,*
   - *If $s_{max,1} \geq s_{max,0}$, it is optimal to choose the user with $x = (s_{max,1}, 1)$.*
   - *If $s_{max,1} < s_{max,0}$, the optimal choice will switch from the user with $x = (s_{max,0}, 0)$ to the user with $x = (s_{max,1}, 1)$ when $s_{max,1}$ increases from 0 to $s_{max,0}$ solely.*

Proof. The first property follows directly from Property 1 and Property 3 of Theorem 1. For the second property, leveraging Property 2 of Theorem 1, we have $\delta^{j,k}(\boldsymbol{x}_2) \leq \delta^{j,k}(\boldsymbol{x}_1) \leq 0$ if $\hat{r}_{1,j} \leq \hat{r}_{2,j}$, $\hat{r}_{1,k} \geq \hat{r}_{2,k}$, and $s_{1,i} = s_{2,i}$ for $1 \leq i \leq N$. Thus, the optimal choice will not be user k in this case. Then, we can conclude that the optimal choice must be in the set $G = \{j\} \cup \{k : \hat{r}_{1,k} < \hat{r}_{2,k}\}$.

For the third property, we have proved in Property 4 of Theorem 1 that $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in s_j if $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$. Hence, $\delta^{j,k}(\boldsymbol{x}_2) \leq \delta^{j,k}(\boldsymbol{x}_1) \leq 0$. As we consider the case of $N = 2$, the optimal choice at state $\boldsymbol{x}_2$ will also be user j. The fourth property can be shown in a similar way by noticing that $\delta^{j,k}(\boldsymbol{x})$ is non-decreasing in s_k when $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$.

For the last property, we recall from Property 5 of Theorem 1 that it is always better to choose the user with a larger s if they are statistically identical and have the same $\hat{r}$. Thus,

we can conclude that the optimal choice must be either the user with $x = (s_{max,1}, 1)$ or the user with $x = (s_{max,0}, 0)$. Without a loss of generality, we assume $x_j = (s_{max,1}, 1)$ and $x_k = (s_{max,0}, 0)$. Now, we distinguish between the following cases

- According to Property 5 of Theorem 1, we can conclude that it is optimal to choose user j when $s_{max,1} \geq s_{max,0}$.
- To determine the optimal choice in the case of $s_{max,1} < s_{max,0}$, we recall that the optimal choice will be user k (i.e., $\delta^{j,k}(\boldsymbol{x}) \geq 0$) if $s_j = 0$ and will be user j (i.e., $\delta^{j,k}(\boldsymbol{x}) \leq 0$) if $s_j = s_k$. At the same time, Property 4 of Theorem 1 tells us that $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in s_j when users j and k are statistically identical. Therefore, we can conclude that the optimal choice will switch from user k to user j when s_j increases from 0 to s_k solely.

□

4. Whittle's Index Policy

Whittle's index policy is a well-known low-complexity heuristic that shows a strong performance in many problems that belong to RMAB [22–24]. In this section, we develop Whittle's index policy for PP. We first present the general procedures we adopt to obtain Whittle's index.

- We first formulate a relaxed version of PP and apply the Lagrangian approach.
- Then, we decouple the problem of minimizing the Lagrangian function into N decoupled problems, each of which only considers a single user. By casting the decoupled problem into an MDP, we investigate the structural properties and performance of the optimal policy.
- Leveraging the results above and under a simple condition, we establish the indexability of the decoupled problem.
- Finally, we obtain the expression of Whittle's index by solving the Bellman equation.

4.1. Relaxed Problem

The first step in obtaining Whittle's index is to formulate the Relaxed Problem (RP). More precisely, instead of requiring the limit on the number of updates allowed per transmission attempt to be met in each time slot, we relax the constraint such that the limit is not violated in an average sense. Then, RP can be formulated as

$$\underset{\phi \in \Phi}{\arg\min} \quad \bar{\Delta}_\phi \triangleq \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N} f_i(s_{i,t})\right) \tag{6a}$$

$$\text{subject to} \quad \bar{\rho}_\phi \triangleq \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N} a_{i,t}\right) \leq M. \tag{6b}$$

As RP is specified, we apply the Lagrangian approach. First of all, we write RP into its Lagrangian form.

$$\mathcal{L}(\lambda, \phi) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N} (f_i(s_{i,t}) + \lambda a_{i,t})\right) - \lambda M,$$

where $\lambda \geq 0$ is the Lagrange multiplier. Then, we investigate the problem of minimizing the Lagrangian function. Since λM is independent of policies, we can ignore it. More precisely, we consider the following minimization problem

$$\underset{\phi \in \Phi}{\text{minimize}} \quad \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_\phi\left(\sum_{t=0}^{T-1}\sum_{i=1}^{N} (f_i(s_{i,t}) + \lambda a_{i,t})\right). \tag{7}$$

4.2. Decoupled Model

In this section, we formulate the decoupled problem and investigate its optimal policy. The decoupled model associated with each user follows the system model with $N = 1$.

Since all the users share the same structure, we drop the user-dependent subscript i for simplicity. Then, the decoupled problem can be formulated as

$$\underset{\phi \in \Phi'}{\text{minimize}} \quad \lim_{T\to\infty} \frac{1}{T}\mathbb{E}_{\phi}\left(\sum_{t=0}^{T-1}(f(s_t)+\lambda a_t)\right), \tag{8}$$

where Φ' is the set of all causal policies when $N=1$. We notice that problem (8) can be cast into the MDP $\mathcal{M}_1(\lambda,-1)$. We define $M=-1$ when there is no restriction on the number of updates allowed per transmission attempt.

We first investigate the structural properties of the optimal policy for $\mathcal{M}_1(\lambda,-1)$ when λ is a given non-negative constant. We start with characterizing the corresponding value function $V(x)$.

Corollary 2 (Extension of Lemma 1)**.** *For $\mathcal{M}_1(\lambda,-1)$, $V(x)$ is non-decreasing in s.*

Proof. The proof follows the same steps as in the proof of Lemma 1. The complete proof can be found in Appendix D. □

Equipped with the above corollary, we can characterize the structural properties of the optimal policy for (8).

Proposition 1 (Optimal policy for decoupled problem)**.** *The optimal policy for the decoupled problem is a threshold policy with the following properties.*

- *The optimal policy can be fully captured by $\boldsymbol{n}=(n_0,n_1)$. More precisely, when the system is at state $(s,\hat{r})$, it is optimal to make a transmission attempt only when $s\geq n_{\hat{r}}$.*
- *$n_0\geq n_1>0$.*

Proof. We define $\Delta V(x)\triangleq V^1(x)-V^0(x)$, where $V^a(x)$ is the value function resulting from taking action a at state x. Then, the optimal action at state x is $a=1$ if $\Delta V(x)<0$, and $a=0$ is optimal otherwise. We use Corollary 2 to characterize the sign of $\Delta V(x)$. The complete proof can be found in Appendix E. □

In the following, we evaluate the performance of the threshold policy detailed in Proposition 1. More precisely, we calculate the expected AoII $\bar{\Delta}_{\boldsymbol{n}}$ and the expected transmission rate $\bar{\rho}_{\boldsymbol{n}}$ resulting from the adoption of threshold policy $\boldsymbol{n}$. We will see in the following that $\bar{\Delta}_{\boldsymbol{n}}$ and $\bar{\rho}_{\boldsymbol{n}}$ are essential for establishing the indexability and obtaining the expression of Whittle's index.

Proposition 2 (Performance)**.** *Under threshold policy $\boldsymbol{n}=(n_0,n_1)$,*

$$\bar{\Delta}_{\boldsymbol{n}}=\pi_0 p\left[\sum_{k=1}^{n_1-1}f(k)(1-p)^{k-1}+(1-p)^{n_1-1}\left(\sum_{k=n_1}^{n_0-1}f(k)c_1^{k-n_1}+c_1^{n_0-n_1}\sum_{k=n_0}^{+\infty}f(k)c_2^{k-n_0}\right)\right],$$

$$\bar{\rho}_{\boldsymbol{n}}=\pi_0 p(1-p)^{n_1-1}\left[\frac{\gamma}{1-c_1}+c_1^{n_0-n_1}\left(\frac{1}{1-c_2}-\frac{\gamma}{1-c_1}\right)\right],$$

where

$$\pi_0=\frac{1}{2+p(1-p)^{n_1-1}\left[\frac{1}{1-c_1}-\frac{1}{p}+c_1^{n_0-n_1}\left(\frac{1}{1-c_2}-\frac{1}{1-c_1}\right)\right]},$$

$c_1=(1-\gamma)(1-p)+\gamma\alpha$, and $c_2=(1-\gamma)\beta+\gamma\alpha$.

Proof. We notice that the dynamic of AoII under the threshold policy can be fully captured by a Discrete-Time Markov Chain (DTMC). Then, combined with the fact that $\hat{r}$ is an independent Bernoulli random variable, we can obtain the desired results from the stationary distribution of the induced DTMC. The complete proof can be found in Appendix F. □

As $f(\cdot)$ can be any non-decreasing function, $\bar{\Delta}$ can grow indefinitely. Thus, it is necessary to require that there exists at least one threshold policy that causes a finite $\bar{\Delta}$. By noting that $1-p \geq c_1 \geq c_2$, we have

$$\begin{aligned}\bar{\Delta} &\geq \pi_0 p\left[\sum_{k=1}^{n_1-1} f(k)c_2^{k-1} + c_2^{n_1-1}\left(\sum_{k=n_1}^{n_0-1} f(k)c_2^{k-n_1} + c_2^{n_0-n_1}\sum_{k=n_0}^{+\infty} f(k)c_2^{k-n_0}\right)\right] \\ &= \pi_0 p\left(\sum_{k=1}^{+\infty} f(k)c_2^{k-1}\right).\end{aligned}$$

The equality is achieved when $n_0 = n_1 = 1$. Then, we can conclude that it is sufficient to require $\sum_{k=1}^{+\infty} f(k)c_2^{k-1} < +\infty$. This will be the underlying assumption throughout the rest of this paper.

4.3. Indexability

In this section, we establish the indexability of the decoupled problem, which ensures the existence of Whittle's index. We start with the definition of indexability.

Definition 2 (Indexability)**.** *The decoupled problem is indexable if the set of states in which $a = 0$ is the optimal action increases with λ, that is,*

$$\lambda' < \lambda \Longrightarrow D(\lambda') \subseteq D(\lambda),$$

where $D(\lambda)$ is the set of states in which $a = 0$ is optimal when Lagrange multiplier λ is adopted.

The Lagrange multiplier λ can be viewed as a cost associated with each transmission attempt. Intuitively, as λ increases, the base station should stay idle (i.e., $a = 0$) for a longer time until s becomes large enough to offset the cost. Although it is intuitively correct that the decoupled problem is indexable, the indexability is hard to establish as the optimal policy is characterized by two thresholds. Thus, Whittle's index does not necessarily exist. However, the indexability can be established when the following condition is satisfied

$$p_{e,i}^0 = 0 \quad for\ 1 \leq i \leq N. \tag{9}$$

Remark 3. *Problem (9) only requires the estimate $\hat{r}_i$ to be perfect when $\hat{r}_i = 0$. In the case of $\hat{r}_i = 1$, we still allow the estimate to be inaccurate.*

When (9) is satisfied, Propositions 1 and 2 reduce to the following

Corollary 3 (Consequences of (9))**.** *When (9) is satisfied, the optimal policy for the decoupled problem (8) is the threshold policy $\boldsymbol{n} = (+\infty, n)$. The corresponding $\bar{\Delta}_{\boldsymbol{n}}$ and $\bar{\rho}_{\boldsymbol{n}}$ are*

$$\bar{\Delta}_{\boldsymbol{n}} = \pi_0 p\left(\sum_{k=1}^{n-1} f(k)(1-p)^{k-1} + (1-p)^{n-1}\sum_{k=n}^{+\infty} f(k)c_1^{k-n}\right),$$

$$\bar{\rho}_{\boldsymbol{n}} = \pi_0 p(1-p)^{n-1}\left(\frac{\gamma}{1-c_1}\right),$$

where

$$\pi_0 = \frac{1}{2 + p(1-p)^{n-1}\left(\frac{1}{1-c_1} - \frac{1}{p}\right)}.$$

Proof. We continue with the same notations as in the proof of Propositions 1 and 2. It is sufficient to show that $n_0 = +\infty$. To this end, we consider the state $x = (s, 0)$. By following the same steps as in the proof of Proposition 1, we have

$$\Delta V(s,0) = \lambda \geq 0.$$

Therefore, it is optimal to stay idle (i.e., $a = 0$) at state $x = (s,0)$ for any $s \geq 0$. Equivalently, $n_0 = +\infty$. Then, the corresponding $\bar{\Delta}_{\boldsymbol{n}}$ and $\bar{\rho}_{\boldsymbol{n}}$ can be calculated as a special case of Proposition 2 where $n_0 = +\infty$, $n_1 = n$, and $p_e^0 = 0$. □

Leveraging Corollary 3, we can establish the indexability of the decoupled problem.

Proposition 3 (Indexability of decoupled problem)**.** *The decoupled problem is indexable when* (9) *is satisfied.*

Proof. According to Proposition 2.2 of [25], we only need to verify that the expected transmission rate $\bar{\rho}_{\boldsymbol{n}}$ is strictly decreasing in n. From Corollary 3, we have

$$\bar{\rho}_{\boldsymbol{n}} = \frac{\gamma\left(\frac{p}{1-c_1}\right)}{\frac{2}{(1-p)^{n-1}} + \left(\frac{p}{1-c_1} - 1\right)}.$$

As $\frac{1}{2} < 1 - p < 1$, we can easily verify that $\bar{\rho}_{\boldsymbol{n}}$ is strictly decreasing in n. Thus, the decoupled problem is indexable when (9) is satisfied. □

4.4. Whittle's Index Policy

In this section, we proceed with finding the expression of Whittle's index and defining Whittle's index policy. First of all, we give the definition of Whittle's index.

Definition 3 (Whittle's index)**.** *When the decoupled problem is indexable, Whittle's index at state x is defined as the infimum λ, such that both actions are equally desirable. Equivalently, Whittle's index at state x is defined as the infimum λ such that $V^0(x) = V^1(x)$.*

Let us denote by W_x the Whittle's index at state x. Then, the expression of Whittle's index is given by the following Proposition.

Proposition 4 (Whittle's index)**.** *When* (9) *is satisfied, Whittle's index is*

$$W_x = \begin{cases} 0 & \text{when } x = (0,\hat{r}) \text{ or } x = (s,0), \\ \dfrac{(1-c_1)\sum\limits_{k=s+1}^{+\infty} f(k)c_1^{k-s-1} - \bar{\Delta}_s}{\frac{(1-c_1)(1-p)-\gamma(1-p-\alpha)}{c_1(1-p-\alpha)}} + \bar{\rho}_s & \text{when } x = (s,1), \end{cases}$$

where $s > 0$ and $c_1 = (1-\gamma)(1-p) + \gamma\alpha$. $\bar{\Delta}_s$ and $\bar{\rho}_s$ are the expected AoII and the expected transmission rate when threshold policy $\boldsymbol{n} = (+\infty, s)$ is adopted, respectively. At the same time, W_x is non-negative and is non-decreasing in s.

Proof. Whittle's indexes at state $x = (0,\hat{r})$ and $x = (s,0)$ are obtained easily from the proof of Proposition 1. For state $x = (s,1)$, we first use backward induction to calculate the expressions of some value functions. Then, the expression of Whittle's index can be obtained from its definition. The complete proof can be found in Appendix G. □

Definition 4 (Whittle's index policy)**.** *At any state $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, the base station will transmit the updates from M users with the largest W_{x_i}. The ties are broken arbitrarily. W_{x_i} is calculated using Proposition 4 with the parameters of user i.*

Remark 4. *Whittle's index policy possesses the structural properties detailed in Corollary 1.*

- *The first two properties can be verified by noting that $W_{x_i} \geq 0$ and the equality holds when $\hat{r}_i = 0$ or $s_i = 0$. At the same time, W_{x_i} is non-decreasing in $\hat{r}_i$.*
- *The third and fourth properties can be verified by noting that W_{x_i} is non-decreasing in s_i.*
- *For the last property, we first notice that $W_{x_j} = W_{x_k}$ when users j and k are statistically identical and $x_j = x_k$. Then, the property can be verified by noting that W_{x_i} is non-decreasing in both s_i and $\hat{r}_i$.*

5. Optimal Policy for Relaxed Problem

In this section, we provide an efficient algorithm to obtain the optimal policy for RP, based on which we will develop another scheduling policy for PP in the next section that is free from indexability. At the same time, the performance of the optimal policy for RP forms a universal lower bound because the following ordering holds

$$\bar{\Delta}_{AoII}^{RP} \leq \bar{\Delta}_{AoII}^{PP},$$

where $\bar{\Delta}_{AoII}^{RP}$ and $\bar{\Delta}_{AoII}^{PP}$ are the minimal expected AoII of RP and PP, respectively.

Remark 5. *Note that the optimal policy for RP may not necessarily be a valid policy for PP, as the transmitter may transmit more than M updates in one transmission attempt under RP-optimal policy.*

To solve RP, we follow the discussion in Section 4.1. More precisely, we take the Lagrangian approach and consider the problem reported in (7). We will see in the following discussion that the optimal policy for RP can be characterized by the optimal policies for problem (7). Therefore, we first cast problem (7) into the MDP $\mathcal{M}_N(\lambda, -1)$. However, the optimal policy for $\mathcal{M}_N(\lambda, -1)$ is difficult to obtain because the state space is infinite. Even though we can make the state space finite by imposing an upper limit on the value of s, the state space and the action space grow exponentially with the number of users in the system. To overcome the difficulty, we investigate the optimal policy for $\mathcal{M}_1^i(\lambda, -1)$ where $1 \leq i \leq N$. The superscript i means that the only user in the system is user i. We will show later that the optimal policy for $\mathcal{M}_N(\lambda, -1)$ can be fully characterized by the optimal policies for $\mathcal{M}_1^i(\lambda, -1)$ where $1 \leq i \leq N$.

5.1. Optimal Policy for Single User

In this section, we tackle the problem of finding the optimal policy for $\mathcal{M}_1^i(\lambda, -1)$. Since the users share the same structure, we ignore the superscript i for simplicity. To find the optimal policy, we first use the Approximating Sequence Method (ASM) introduced in [26] to make the state space finite. More precisely, we impose $s \leq m$ where m is a predetermined upper limit. The state transition probabilities $P'_{s,s'}(a, \hat{r})$ are modified in the following way

$$P'_{s,s'}(a, \hat{r}) = \begin{cases} P_{s,s'}(a, \hat{r}) & if \ \ s' < m, \\ P_{s,s'}(a, \hat{r}) + \sum_{z>m} P_{s,z}(a, \hat{r}) & if \ \ s' = m. \end{cases} \tag{10}$$

The action space and the instant cost remain unchanged. Then, we can apply Relative Value Iteration (RVI) with convergence criteria ϵ to obtain the optimal policy. We notice that $\mathcal{M}_1(\lambda, -1)$ coincides with the decoupled model studied in Section 4.2. Hence, we can utilize the threshold structure of the optimal policy to improve RVI. To this end, we class a state as active if the optimal action at this state is $a = 1$. Then, the threshold structure detailed in Proposition 1 tells us the following. For any state x, if there exists an active state x_1 with $s_1 \leq s$ and $\hat{r}_1 \leq \hat{r}$, then x must also be active. Hence, we can determine the optimal action at state x immediately instead of comparing all feasible actions. In this way, we can reduce the running time of RVI. The pseudocode for the improved RVI can be found in Algorithm A1 of Appendix M. A similar technique is also presented in [5].

For $\mathcal{M}_1(\lambda,-1)$, when problem (9) is satisfied, Whittle's index exists and can be calculated efficiently using Proposition 4. Therefore, we can obtain the optimal policy using Whittle's index and further reduce the computational complexity. To this end, we denote by $\boldsymbol{n}_\lambda$ the optimal policy for $\mathcal{M}_1(\lambda,-1)$ and present the following proposition

Proposition 5 (Optimal deterministic policy)**.** *When (9) is satisfied, the optimal policy for $\mathcal{M}_1(\lambda,-1)$ is $\boldsymbol{n}_\lambda = (+\infty, n)$ where n is given by*

$$n = \begin{cases} 1 & if\ \lambda = 0, \\ \max\{s \in \mathbb{N}_0 : W_s \leq \lambda\} + 1 & if\ \lambda > 0. \end{cases}$$

W_s is the Whittle's index at state $(s,1)$.

Proof. We first notice that $\mathcal{M}_1(\lambda,-1)$ coincides with the decoupled model studied in Section 4.2. Then, we show the optimal action for each state with $\hat{r}=1$ using the definition of Whittle's index and the fact that the decoupled problem is indexable when (9) is satisfied. The complete proof can be found in Appendix H. □

In the following, we provide a randomized policy that is also optimal for $\mathcal{M}_1(\lambda,-1)$. We will see later that the randomized policy is the key to obtaining the optimal policy for RP.

Theorem 2 (Optimal randomized policy)**.** *There exist two deterministic policies $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$, which are both optimal for $\mathcal{M}_1(\lambda,-1)$. We consider the following randomized policy $\boldsymbol{n}_\lambda$: every time the system reaches state $(0,0)$, the base station will make the choice between $\boldsymbol{n}_{\lambda_-}$ with probability μ and $\boldsymbol{n}_{\lambda_+}$ with probability $1-\mu$. The chosen policy will be followed until the next choice. Then, the randomized policy $\boldsymbol{n}_\lambda$ is optimal for $\mathcal{M}_1(\lambda,-1)$ under any $\mu \in [0,1]$.*

Proof. We show that our system verifies the assumptions given in [27]. Then, leveraging the characteristics of our system, we can obtain the optimal randomized policy. The complete proof can be found in Appendix I. □

In practice, we approximate $\lambda_+ \approx \lambda + \xi$ and $\lambda_- \approx \lambda - \xi$ where ξ is a small perturbation. Then, the deterministic policies $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$ can be obtained by following the discussion at the beginning of this subsection. Note that, in most cases, $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$ are the same.

5.2. Optimal Policy for RP

In this section, we characterize the optimal policy for RP. Let us denote by $V(\boldsymbol{x})$ and $V^i(x_i)$ the value functions of $\mathcal{M}_N(\lambda,-1)$ and $\mathcal{M}_1^i(\lambda,-1)$, respectively. Then, we can prove the following

Proposition 6 (Separability)**.** $V(\boldsymbol{x}) = \sum_{i=1}^N V^i(x_i)$ *where* $\boldsymbol{x} = (x_1,\ldots,x_N)$*. In other words, the policy, under which each user adopts its own optimal policy, is optimal for $\mathcal{M}_N(\lambda,-1)$.*

Proof. We show $V(\boldsymbol{x}) = \sum_{i=1}^N V^i(x_i)$ by comparing the Bellman equations they must satisfy. The complete proof can be found in Appendix J. □

We denote the optimal policy for $\mathcal{M}_N(\lambda,-1)$ as $\phi_\lambda = [\boldsymbol{n}_{\lambda,1},\ldots,\boldsymbol{n}_{\lambda,N}]$ where $\boldsymbol{n}_{\lambda,i}$ is the optimal policy for $\mathcal{M}_1^i(\lambda,-1)$. For simplicity, we define $\bar{\Delta}(\lambda)$ and $\bar{\rho}(\lambda)$ as the expected AoII and the expected transmission rate associated with ϕ_λ, respectively. $\bar{\Delta}^i(\lambda)$ and $\bar{\rho}^i(\lambda)$ are defined analogously for user i under policy $\boldsymbol{n}_{\lambda,i}$. We also define $\lambda^* \triangleq \inf\{\lambda > 0 : \bar{\rho}(\lambda) \leq M\}$. With Proposition 6 and the above definitions in mind, we proceed with constructing the optimal policy for RP.

Theorem 3 (Optimal policy for RP). *The optimal policy for RP can be characterized by two deterministic policies* $\phi_{\lambda_+^*} = [\boldsymbol{n}_{\lambda_+^*,1}, \ldots, \boldsymbol{n}_{\lambda_+^*,N}]$ *and* $\phi_{\lambda_-^*} = [\boldsymbol{n}_{\lambda_-^*,1}, \ldots, \boldsymbol{n}_{\lambda_-^*,N}]$ *where* $\boldsymbol{n}_{\lambda_+^*,i}$ *and* $\boldsymbol{n}_{\lambda_-^*,i}$ *are both the optimal deterministic policies for* $\mathcal{M}_1^i(\lambda^*, -1)$. *Then, we mix* $\phi_{\lambda_+^*}$ *and* $\phi_{\lambda_-^*}$ *in the following way: for each user* i, *every time the user reaches state* $(0,0)$, *the base station will make the choice between* $\boldsymbol{n}_{\lambda_-^*,i}$ *with probability* μ_i *and* $\boldsymbol{n}_{\lambda_+^*,i}$ *with probability* $1-\mu_i$. *The chosen policy will be followed by user* i *until the next choice. Where* $1 \le i \le N$, *the* μ_i *is chosen in such a way as to satisfy*

$$\sum_{i=1}^{N} \bar{\rho}^i(\lambda^*) = \sum_{i=1}^{N} \left(\mu_i \bar{\rho}^i(\lambda_-^*) + (1-\mu_i)\bar{\rho}^i(\lambda_+^*) \right) = M. \tag{11}$$

Then, the mixed policy, denoted by ϕ_{λ^*}, *is optimal for RP.*

Proof. According to Lemma 3.10 of [27], a policy is optimal for RP if

1. It is optimal for $\mathcal{M}_N(\lambda^*, -1)$;
2. The resulting expected transmission rate is equal to M.

Then, we construct such a policy using Theorem 2 and Proposition 6. The complete proof can be found in Appendix K. □

Since we approximate $\lambda_+^* \approx \lambda^* + \xi$ and $\lambda_-^* \approx \lambda^* - \xi$ in practice, $\bar{\rho}^i(\lambda_+^*) \le \bar{\rho}^i(\lambda_-^*)$ for all i according to the monotonicity given by Lemma 3.4 of [27]. Combining with the definition of λ^*, we must have $\bar{\rho}(\lambda_+^*) \le M < \bar{\rho}(\lambda_-^*)$. Therefore, we can always find μ_i's that realize (11). In this paper, we choose

$$\mu_i = \mu = \frac{M - \bar{\rho}(\lambda_+^*)}{\bar{\rho}(\lambda_-^*) - \bar{\rho}(\lambda_+^*)}, \quad \text{for } 1 \le i \le N. \tag{12}$$

Then, we describe the algorithm used to obtain the optimal policy for RP. As detailed in Theorem 3, it is essential to find λ^*. To this end, we recall that, for any user i under given λ, the optimal deterministic policy $\boldsymbol{n}_{\lambda,i}$ can be obtained using the results in Section 5.1 and the resulting expected transmission rate $\bar{\rho}^i(\lambda)$ is given by Proposition 2. Since $\bar{\rho}^i(\lambda)$ is non-increasing in λ for all i according to Lemma 3.4 of [27], $\bar{\rho}(\lambda) = \sum_{i=1}^{N} \bar{\rho}^i(\lambda)$ is also non-increasing in λ. Hence, we can regard $\bar{\rho}(\lambda)$ as a non-increasing function of λ. Then, according to the definition of λ^*, we can use the Bisection search to obtain λ^* efficiently. The main steps can be summarized as follows.

1. Initialize $\lambda_- = 0$ and $\lambda_+ = 1$.
2. Do $\lambda_- = \lambda_+$ and $\lambda_+ = 2\lambda_+$ until $\bar{\rho}(\lambda_+) < M$.
3. Run Bisection search on the interval $[\lambda_-, \lambda_+]$ until the tolerance 2ξ is met.

Then, λ_-^* and λ_+^* can simply be the boundaries of the final interval. The pseudocode for the Bisection search can be found in Algorithm A2 of Appendix M. After obtaining λ_-^* and λ_+^*, the optimal policy ϕ_{λ^*} is detailed in Theorem 3 and the mixing probabilities μ_i's are given by (12).

Remark 6. *We recall that the optimal deterministic policy for each user can be characterized by two positive thresholds (i.e.,* $n_0, n_1 > 0$*). Consequently, under RP-optimal policy, the base station will never choose the user at state* $(0, \hat{r})$. *Then, when* M *increases, the expected transmission rate achieved by RP-optimal policy will saturate before* M *reaches* N. *When the expected transmission rate saturates, the RP-optimal policy is* $\phi^* = [\boldsymbol{n}_1, \ldots, \boldsymbol{n}_N]$ *where* $\boldsymbol{n}_i = (1,1)$ *for* $1 \le i \le N$. *The saturation happens when* M *is larger than or equal to the expected transmission rate achieved by* ϕ^*.

6. Indexed Priority Policy

Although the performance of Whittle's index policy is known to be good, it requires indexability, which is usually difficult to establish. In this section, based on the primal-dual heuristic introduced in [28], we develop a policy that does not require indexability

and has comparable performance to Whittle's index policy. We start with presenting the primal-dual heuristic.

6.1. Primal-Dual Heuristic

The heuristic is based on the optimal primal and dual solution pair to the linear program associated with RP. To introduce the linear program, we define $\pi_{x_i}^{a_i}(\phi) \geq 0$ as the expected time that user i is at state x_i and action a_i is taken according to policy ϕ. Then, for any ϕ, $\pi_{x_i}^{a_i}(\phi)$ must satisfy the following problems

$$\pi_{x_i}^0(\phi) + \pi_{x_i}^1(\phi) = \sum_{x_i'} \sum_{a_i'} P_{x_i', x_i}(a_i') \pi_{x_i'}^{a_i'}(\phi), \quad \forall x_i, i.$$

$$\sum_{x_i} \sum_{a_i} \pi_{x_i}^{a_i}(\phi) = 1, \quad \forall i.$$

The objective function of RP can be rewritten as

$$\underset{\phi \in \Phi}{\text{minimize}} \quad \sum_{i=1}^{N} \sum_{x_i, a_i} C(x_i) \pi_{x_i}^{a_i}(\phi),$$

where $C(x_i) = f_i(s_i)$ is the instant cost at state x_i. The constraint on the expected transmission rate can be rewritten as

$$\sum_{i=1}^{N} \sum_{x_i} \pi_{x_i}^1(\phi) \leq M.$$

Thus, the linear program associated with RP can be formulated as the following

$$\underset{\pi_{x_i}^{a_i}}{\text{minimize}} \quad \sum_{i=1}^{N} \sum_{x_i, a_i} C(x_i) \pi_{x_i}^{a_i} \tag{13a}$$

$$\text{subject to} \quad \pi_{x_i}^0 + \pi_{x_i}^1 - \sum_{x_i'} \sum_{a_i'} P_{x_i', x_i}(a_i') \pi_{x_i'}^{a_i'} = 0, \quad \forall x_i, i, \tag{13b}$$

$$\sum_{x_i} \sum_{a_i} \pi_{x_i}^{a_i} = 1, \quad \forall i, \tag{13c}$$

$$\sum_{i=1}^{N} \sum_{x_i} \pi_{x_i}^1 \leq M, \tag{13d}$$

$$\pi_{x_i}^{a_i} \geq 0, \quad \forall x_i, a_i, i. \tag{13e}$$

The corresponding dual problem is

$$\underset{\sigma, \sigma_i, \sigma_{x_i}}{\text{maximize}} \quad \sum_{i=1}^{N} \sigma_i - M\sigma \tag{14a}$$

$$\text{subject to} \quad \sigma_{x_i} + \sigma_i - \sum_{x_i'} P_{x_i, x_i'}(0) \sigma_{x_i'} \leq C(x_i), \quad \forall x_i, i, \tag{14b}$$

$$\sigma_{x_i} + \sigma_i - \sum_{x_i'} P_{x_i, x_i'}(1) \sigma_{x_i'} - \sigma \leq C(x_i), \quad \forall x_i, i, \tag{14c}$$

$$\sigma \geq 0. \tag{14d}$$

Let $\{\bar{\pi}_{x_i}^{a_i}\}$ and $\{\bar{\sigma}, \bar{\sigma}_i, \bar{\sigma}_{x_i}\}$ be the optimal primal and dual solution pair to the problems reported in (13) and (14). We define

$$\bar{\psi}_{x_i}^0 = \sum_{x_i'} P_{x_i, x_i'}(0) \bar{\sigma}_{x_i'} + C(x_i) - \bar{\sigma}_i - \bar{\sigma}_{x_i} \geq 0,$$

$$\bar{\psi}_{x_i}^1 = \sum_{x_i'} P_{x_i, x_i'}(1) \bar{\sigma}_{x_i'} + \bar{\sigma} + C(x_i) - \bar{\sigma}_i - \bar{\sigma}_{x_i} \geq 0.$$

For any state $\boldsymbol{x} = (x_1, \ldots, x_N)$, let $h(\boldsymbol{x}) = \sum_{i=1}^{N} \mathbb{1}_{\{\pi^1_{x_i} > 0\}}$. Then, the heuristic operates in the following way

- If $h(\boldsymbol{x}) \geq M$, the base station will choose the M users with the largest $\bar{\psi}^0_{x_i}$ among the $h(\boldsymbol{x})$ users.
- If $h(\boldsymbol{x}) < M$, these $h(\boldsymbol{x})$ users are chosen by the base station. The base station will choose $M - h(\boldsymbol{x})$ additional users with the smallest $\bar{\psi}^1_{x_i}$.

However, Linear Programming (LP) is a very general technique and does not appear to take advantage of the special structure of the problem. Although there are algorithms for solving rational LP that take time polynomial in the number of variables and constraints, they run extremely slowly in practice [29]. For our problem, we notice that the users have separate activity areas that are linked through a common resource constraint. Therefore, the primal problem can be solved using Dantzig-Wolfe decomposition. Even so, the problem is still computationally demanding when the system scales up. We recall that we solved the exact problem efficiently using MDP-specific algorithms in Section 5. It is more efficient because of the following reasons

- According to Proposition 6, we can decompose the problem into N subproblems.
- For each subproblem, the threshold structure of the optimal policy is utilized to reduce the running time of RVI.
- As we will see later, the developed policy can be obtained directly from the result of RVI in practice.

In the following, we will translate the results in Section 5 into the optimal primal and dual solution pair and propose Indexed priority policy.

6.2. Indexed Priority Policy

We first define the Lagrangian function associated with (13).

$$\mathcal{L}(\pi^{a_i}_{x_i}, \sigma, \sigma_i, \sigma_{x_i}, \psi^{a_i}_{x_i}) = \left(\sum_{i=1}^{N} \sum_{x_i, a_i} C(x_i) \pi^{a_i}_{x_i} \right) + \sum_{i, x_i} \sigma_{x_i} \left(\sum_{x'_i} \sum_{a'_i} P_{x'_i, x_i}(a'_i) \pi^{a'_i}_{x'_i} - \pi^0_{x_i} - \pi^1_{x_i} \right) + \sum_{i=1}^{N} \sigma_i \left(1 - \sum_{x_i} \sum_{a_i} \pi^{a_i}_{x_i} \right) + \sigma \left(\sum_{i=1}^{N} \sum_{x_i} \pi^1_{x_i} - M \right) - \sum_{i, x_i, a_i} \psi^{a_i}_{x_i} \pi^{a_i}_{x_i}.$$

Then, the corresponding Lagrangian dual function is

$$g(\sigma, \sigma_i, \sigma_{x_i}, \psi^{a_i}_{x_i}) = \inf_{\pi^{a_i}_{x_i}} \mathcal{L}(\pi^{a_i}_{x_i}, \sigma, \sigma_i, \sigma_{x_i}, \psi^{a_i}_{x_i}).$$

Let π_{x_i} be the expected time that user i is at state x_i caused by the adoption of ϕ_{λ^*}, where ϕ_{λ^*} is the optimal policy detailed in Theorem 3. Then, we define $\{\pi^{a_i}_{x_i}\}$ as follows

- State x_i is where randomization happens (randomization happens when the actions suggested by the two optimal deterministic policies are different), and it has a value of $\pi^0_{x_i} = a_{\boldsymbol{n}_{\lambda^*_-, i}}(x_i)(1 - \mu_i)\pi_{x_i} + a_{\boldsymbol{n}_{\lambda^*_+, i}}(x_i)\mu_i \pi_{x_i}$ and $\pi^1_{x_i} = \pi_{x_i} - \pi^0_{x_i}$ where μ_i is given by (12) and $a_{\boldsymbol{n}_{\lambda,i}}(x_i)$ is the action suggested by $\boldsymbol{n}_{\lambda,i}$ at state x_i.
- For other values of x_i, we have $\pi^0_{x_i} = (1 - a_{\boldsymbol{n}_{\lambda^*, i}}(x_i))\pi_{x_i}$ and $\pi^1_{x_i} = \pi_{x_i} - \pi^0_{x_i}$.

We also define $\sigma = \lambda^*$, $\sigma_i = \theta_i$, and $\sigma_{x_i} = V^i(x_i)$ where λ^* is specified in Section 5.2, θ_i is the optimal value of $\mathcal{M}^i_1(\lambda^*, -1)$, and $V^i(x_i)$ is the value function associated with $\mathcal{M}^i_1(\lambda^*, -1)$. Lastly, we define $\{\psi^{a_i}_{x_i}\}$ as follows

$$\psi^0_{x_i} = \sum_{x'_i} P_{x_i, x'_i}(0)\sigma_{x'_i} + C(x_i) - \sigma_i - \sigma_{x_i},$$

$$\psi^1_{x_i} = \sum_{x'_i} P_{x_i, x'_i}(1)\sigma_{x'_i} + \sigma + C(x_i) - \sigma_i - \sigma_{x_i}.$$

Then, we can prove the following proposition.

Proposition 7 (Optimal solution pair)**.** *$\{\pi_{x_i}^{a_i}\}$ and $\{\sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i}\}$ are primal and dual solutions to (13), respectively.*

Proof. Since (13) is linear and strictly feasible, it is sufficient to show that $\{\pi_{x_i}^{a_i}\}$ and $\{\sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i}\}$ verify the KKT conditions, which can be expressed as the following four conditions.

1. Primal feasibility: the constraints in (13) are satisfied.
2. Dual feasibility: $\sigma \geq 0$ and $\psi_{x_i}^{a_i} \geq 0$ for all x_i, a_i, and i.
3. Complementary slackness: $\sigma\big(\sum_{i=1}^{N}\sum_{x_i}\pi_{x_i}^{1} - M\big) = 0$ and $\psi_{x_i}^{a_i}\pi_{x_i}^{a_i} = 0$ for all x_i, a_i, and i.
4. Stationarity: the gradient of $\mathcal{L}(\pi_{x_i}^{a_i}, \sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i})$ with respect to $\{\pi_{x_i}^{a_i}\}$ vanishes.

Apparently, the first condition is satisfied by $\{\pi_{x_i}^{a_i}\}$. For the second condition, $\sigma \geq 0$ since $\sigma = \lambda^* \geq 0$ by definition. For $\psi_{x_i}^{a_i}$, we can verify that $\psi_{x_i}^{a_i} = V^{i,a_i}(x_i) - V^i(x_i)$ where $V^{i,a_i}(x_i)$ is the value function resulting from taking action a_i at state x_i. Then, the non-negativity is guaranteed by the Bellman equation. For the third condition, the first term is zero because we choose the μ_i's given by (12). For the second term, we recall that $\psi_{x_i}^{a_i} = V^{i,a_i}(x_i) - V^i(x_i)$. According to the definition of $\pi_{x_i}^{a_i}$, we know $V^i(x_i) = V^{i,a_i}(x_i)$ if $\pi_{x_i}^{a_i} > 0$. Combined together, we can conclude that $\psi_{x_i}^{a_i} = 0$ when $\pi_{x_i}^{a_i} > 0$. Thus, the third condition is satisfied. For the last condition, setting the gradient equal to zero yields a system of linear equations. More precisely, for each x_i and $1 \leq i \leq N$

$$\begin{cases} \sum\limits_{x_i'} P_{x_i,x_i'}(0)\sigma_{x_i'} + C(x_i) = \sigma_{x_i} + \sigma_i + \psi_{x_i}^{0}. \\ \sum\limits_{x_i'} P_{x_i,x_i'}(1)\sigma_{x_i'} + \sigma + C(x_i) = \sigma_{x_i} + \sigma_i + \psi_{x_i}^{1}. \end{cases}$$

Then, $\{\sigma, \sigma_i, \sigma_{x_i}, \psi_{x_i}^{a_i}\}$ verifies the system of linear equations by definition. Since all four conditions are satisfied, we can conclude our proof. □

According to Proposition 7, we know that $\{\pi_{x_i}^{a_i}\}$ and $\{\sigma, \sigma_i, \sigma_{x_i}\}$ defined above are the optimal solutions to problems (13) and (14), respectively. As the optimal solutions are obtained, we can adopt the heuristic detailed in Section 6.1.

The heuristic can be expressed equivalently as an index policy. To this end, we define the index I_{x_i} for state x_i as

$$I_{x_i} \triangleq \bar{\psi}_{x_i}^{0} - \bar{\psi}_{x_i}^{1}.$$

According to the complementary slackness, I_{x_i} can be reduced to the following.

- For state x_i such that $\bar{\pi}_{x_i}^{1} > 0$ and $\bar{\pi}_{x_i}^{0} = 0$, we have $\bar{\psi}_{x_i}^{1} = 0$. Therefore, $I_{x_i} = \bar{\psi}_{x_i}^{0} \geq 0$.
- For state x_i such that $\bar{\pi}_{x_i}^{1} > 0$ and $\bar{\pi}_{x_i}^{0} > 0$, we have $\bar{\psi}_{x_i}^{1} = \bar{\psi}_{x_i}^{0} = 0$. Therefore, $I_{x_i} = 0$.
- For state x_i such that $\bar{\pi}_{x_i}^{1} = 0$ and $\bar{\pi}_{x_i}^{0} > 0$, we have $\bar{\psi}_{x_i}^{0} = 0$. Therefore, $I_{x_i} = -\bar{\psi}_{x_i}^{1} \leq 0$.

We can show that I_{x_i} possesses the following properties.

Proposition 8 (Properties of I_{x_i})**.** *For $1 \leq i \leq N$, $I_{x_i} \geq -\lambda^*$ for any x_i. The equality holds when $\hat{r}_i = p_{e,i}^{0} = 0$ or $s_i = 0$. At the same time, I_{x_i} is non-decreasing in both s_i and $\hat{r}_i$.*

Proof. We notice that I_{x_i} can be expressed as a function of $V^i(x_i)$ and λ^*. Meanwhile, $\mathcal{M}_1^i(\lambda^*, -1)$ coincides with the decoupled model studied in Section 4.2. Then, we can verify the properties of I_{x_i} using the results in Section 4.2. The complete proof can be found in Appendix L. □

Comparing with the heuristic detailed in Section 6.1, we can define the Indexed priority policy.

Definition 5 (Indexed priority policy)**.** *At any state $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, the base station will transmit the updates from M users with the largest I_{x_i}. The ties are broken arbitrarily.*

Remark 7. *Indexed priority policy belongs to the class of priority policies introduced in [30]. These priority policies are asymptotically optimal when certain conditions are satisfied.*

Remark 8. *Indexed priority policy possesses the structural properties detailed in Corollary 1.*

- *The first two properties can be verified by noting that $I_{x_i} \geq -\lambda^*$ and the equality holds when $\hat{r}_i = p^0_{e,i} = 0$ or $s_i = 0$. At the same time, I_{x_i} is non-decreasing in $\hat{r}_i$.*
- *The third and fourth properties can be verified by noting that I_{x_i} is non-decreasing in s_i.*
- *For the last property, we first notice that $I_{x_j} = I_{x_k}$ when users j and k are statistically identical and $x_j = x_k$. Then, the property can be verified by noting that I_{x_i} is non-decreasing in both s_i and $\hat{r}_i$.*

We notice that θ_i's and $C(x_i)$'s are canceled out by the definition of I_{x_i}. Therefore, I_{x_i} can be calculated using λ^* and the value function of $\mathcal{M}^i_1(\lambda^*, -1)$. In practice, we can use either λ^*_- or λ^*_+ to approximate λ^*, and the value function can be approximated by the result of the RVI detailed in Section 5.1. Since the state space is infinite, we only calculate a finite number of $V^i(x_i)$, the number of which depends on the truncation parameter m of ASM. Meanwhile, the probabilities $P_{x_i,x'_i}(a_i)$ in I_{x_i} are modified according to (10).

7. Numerical Results

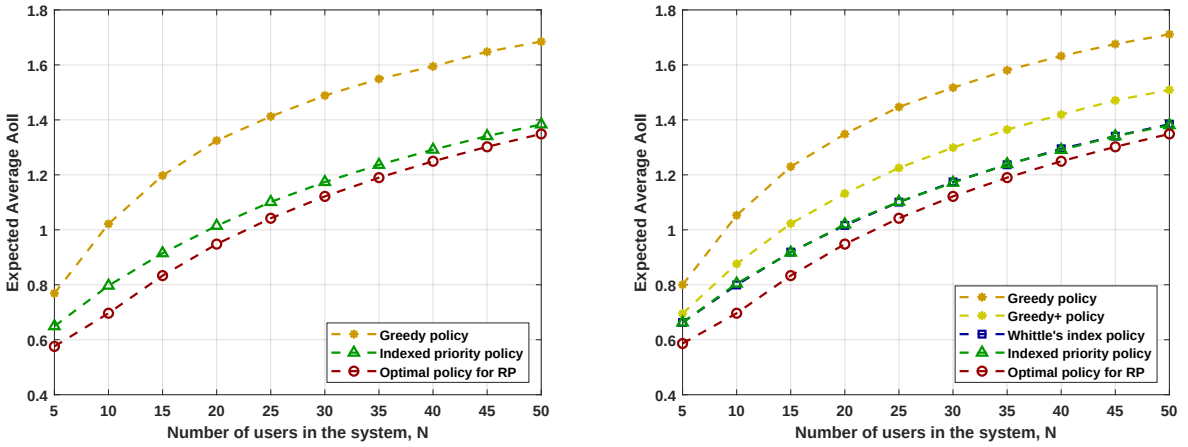
In this section, we provide numerical results to showcase the performance of the developed scheduling policies. To eliminate the effect of N, we plot the expected average AoII. In particular, we provide the expected average AoII achieved by the Indexed priority policy and Whittle's index policy when $M = 1$. The policies are calculated using the results detailed in Sections 4–6. When obtaining the Indexed priority policy, we set the tolerance in the Bisection search to $\xi = 0.005$. Meanwhile, we choose the truncation parameter in ASM $m = 800$ and the convergence criteria in RVI $\epsilon = 0.01$. We notice that the calculation of Whittle's index involves an infinite sum. In practice, we approximate the result by replacing $+\infty$ with a large enough number k_{max}. Here, we choose $k_{max} = 800$. For both scheduling policies, the resulting expected average AoII is obtained via simulations. Each data point is the average of 15 runs with 15,000 time slots considered in each run.

We also compare the developed policies with the optimal policy for RP, which can be calculated by following the discussion in Section 5.2. We adopt the same choices of parameters as we used to obtain the developed policies. The corresponding performance is calculated using Proposition 2. Like before, the infinite sum is approximated by replacing $+\infty$ with $k_{max} = 800$. We also provide the expected average AoII achieved by the Greedy policy to show the performance advantages of the developed policies. When the Greedy policy is adopted, the base station always chooses the user with the largest AoII. The resulting expected average AoII is obtained via the same simulations as applied to the developed policies.

Figures 3 and 4 illustrate the performance when the source processes have different dynamics and when each user's communication goal is different, respectively. Figure 3a provides the performance when $p_i = 0.05 + \frac{0.4(i-1)}{N-1}$ for $1 \leq i \leq N$. For other parameters, the users make the same choices. More precisely, $f_i(s) = s$, $\gamma_i = 0.6$, and $p^0_{e,i} = p^1_{e,i} = 0.1$ for $1 \leq i \leq N$. Figure 4a provides the performance when $f_i(s) = s^{0.5+\frac{i-1}{N-1}}$ for $1 \leq i \leq N$. Same as before, the users make the same choices for other parameters. More precisely, $p_i = 0.3$, $\gamma_i = 0.6$, and $p^0_{e,i} = p^1_{e,i} = 0.1$ for $1 \leq i \leq N$. In Figures 3b and 4b, we force $p^0_{e,i} = 0$ for all users to ensure the existence of Whittle's index. Other choices remain the same as in Figures 3a and 4a. According to Corollary 1, the optimal policy will never choose the user with $\hat{r} = p^0_e = 0$ unless it is to break the tie. Therefore, in Figures 3b and 4b, we also consider the Greedy+ policy where the base station always chooses the user

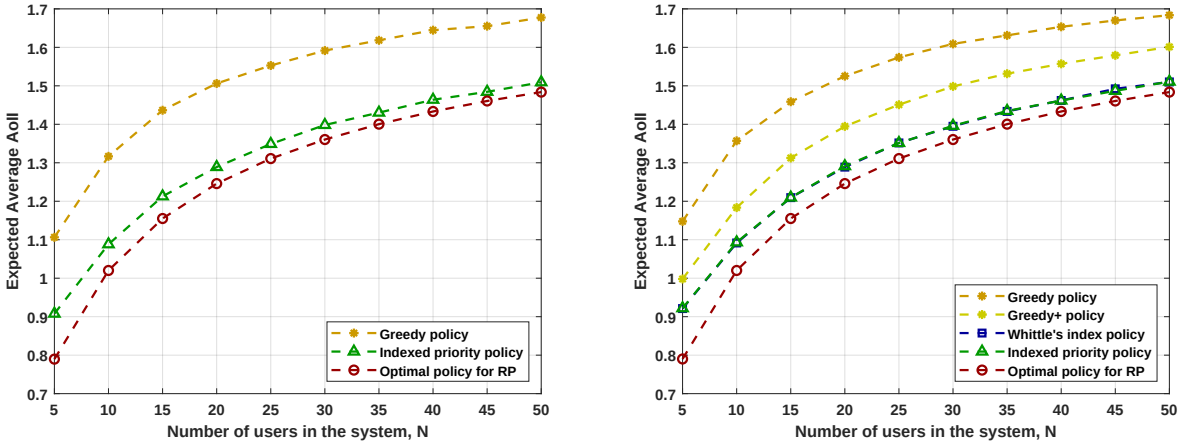
with the largest AoII among the users with $\hat{r} = 1$. The resulting expected average AoII is obtained via the same simulations as applied to the Greedy policy.

Figure 5 shows the performance in systems where the parameters for each user are generated uniformly and randomly within their ranges. In Figure 5a, we consider $N = 5$, $\gamma \in [0, 1]$, $p \in [0.05, 0.45]$, $p_e^{\hat{r}} \in [0, 0.45]$, and $f(s) = s^\tau$, where $\tau \in [0.5, 1.5]$. There are a total of 300 different choices and the results are sorted by the performance of RP-optimal policy in ascending order. Figure 5b adopts the same system settings except that we impose $p_{e,i}^0 = 0$ for $1 \leq i \leq N$ to ensure the feasibility of Whittle's index policy. Meanwhile, we ignore the Greedy policy since the Greedy+ policy achieves a better performance, as indicated by Figures 3b and 4b.



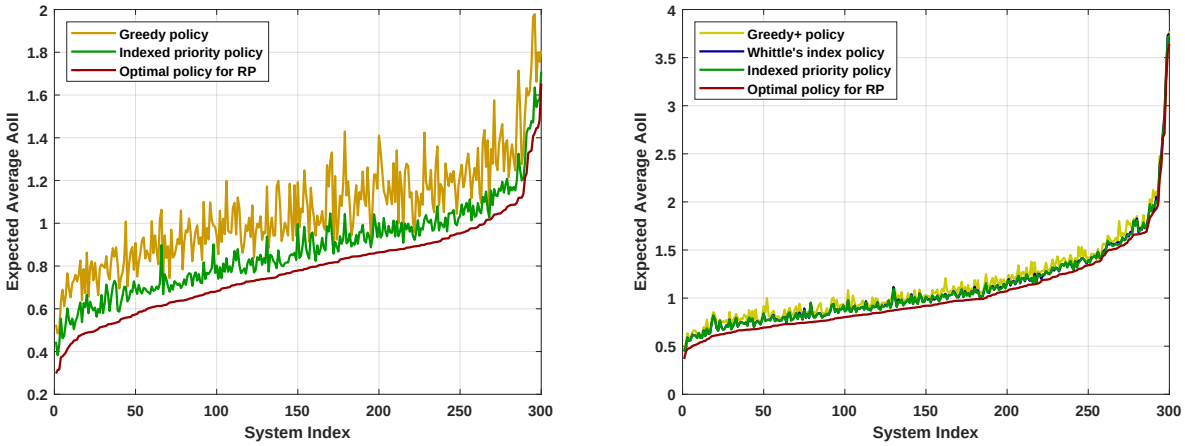
(**a**) When $p_e^0 = 0.1$. (**b**) When $p_e^0 = 0$.

Figure 3. Performance when the source processes vary. We choose $p_i = 0.05 + \frac{0.4(i-1)}{N-1}$, $f_i(s) = s$, $\gamma_i = 0.6$, $p_{e,i}^0 = p_e^0$, and $p_{e,i}^1 = 0.1$ for $1 \leq i \leq N$.



(**a**) When $p_e^0 = 0.1$. (**b**) When $p_e^0 = 0$.

Figure 4. Performance when the communication goals vary. We choose $f_i(s) = s^{0.5+\frac{i-1}{N-1}}$, $p_i = 0.3$, $\gamma_i = 0.6$, $p_{e,i}^0 = p_e^0$, and $p_{e,i}^1 = 0.1$ for $1 \leq i \leq N$.



(**a**) When $I = [0, 0.45]$. (**b**) When $I = \{0\}$.

Figure 5. Performance in systems with random parameters when $N = 5$. The parameters for each user are chosen randomly within the following intervals: $\gamma \in [0, 1]$, $p \in [0.05, 0.45]$, $p_e^0 \in I$, $p_e^1 \in [0, 0.45]$, and $f(s) = s^\tau$ where $\tau \in [0.5, 1.5]$.

We can make the following observations from the figures.

- The Greedy+ policy yields a smaller expected average AoII than that achieved by the Greedy policy. Recall that we obtained the Greedy+ policy by applying the structural properties detailed in Corollary 1. Therefore, simple applications of the structural properties of the optimal policy can improve the performance of scheduling policies.
- The Indexed priority policy has comparable performance to Whittle's index policy in all the system settings considered. The two policies have their own advantages. The Indexed priority policy has a broader scope of application, while Whittle's index policy has a lower computational complexity.
- The performance of the Indexed priority policy and Whittle's index policy is better than that of the Greedy/Greedy+ policies and is not far from the performance of the RP-optimal policy. Recall that the performance of the RP-optimal policy forms a universal lower bound on the performance of all admissible policies for PP. Hence, we can conclude that both the Indexed priority policy and Whittle's index policy achieve good performances.

8. Conclusions

In this paper, we studied the problem of minimizing the Age of Incorrect Information in a slotted-time system where a base station needs to schedule M users among N available users. Meanwhile, the base station has access to imperfect channel state information in each time slot. The problem is a restless multi-armed bandit problem which is SPACE-hard. However, by casting the problem into a Markov decision process, we obtain the structural properties of the optimal policy. Then, we introduce a relaxed version of the original problem and investigate the decoupled model. Under a simple condition, we establish the indexability of the decoupled problem and obtain the expression of Whittle's index. On this basis, we developed Whittle's index policy. To get rid of the requirement for indexability, we developed the Indexed priority policy based on the optimal policy for the relaxed problem. The characteristics of the relaxed problem are explored to make the calculation of its optimal policy more efficient. Finally, through numerical results, we show that simple applications of the structural properties can improve the performance of scheduling policies. Moreover, Whittle's index policy and the Indexed priority policy achieve good and comparable performances.

Author Contributions: Formal analysis, Y.C.; Investigation, Y.C.; Methodology, Y.C.; Supervision, A.E.; Validation, Y.C.; Writing—original draft, Y.C.; Writing—review & editing, Y.C. and A.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Lemma 1

We consider two states, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, that differ only in the value of s_j. Without the loss of generality, we assume $s_{1,j} < s_{2,j}$. Then, it is sufficient to show that, for any $1 \leq j \leq N$, $V(\boldsymbol{x}_1) \leq V(\boldsymbol{x}_2)$. Leveraging the iterative nature of VIA, we use mathematical induction to prove the monotonicity. First of all, the base case (i.e., $\nu = 0$) is true by initialization. We assume the lemma holds at iteration ν. Then, we want to examine whether it holds at iteration $\nu + 1$. The update step reported in problem (5) can be rewritten as follows.

$$V_{\nu+1}(\boldsymbol{x}) = \min_{\boldsymbol{a} \in \mathcal{A}_N(1)} V^{\boldsymbol{a}}_{\nu+1}(\boldsymbol{x}), \tag{A1}$$

where

$$V^{\boldsymbol{a}}_{\nu+1}(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}' - \{x'_j\}} \left\{ \left(\prod_{i \neq j} P_{x_i, x'_i}(a_i) \right) \sum_{\hat{r}'_j} P(\hat{r}'_j) U^j_\nu(\boldsymbol{x}, \boldsymbol{x}') \right\},$$

$$U^j_\nu(\boldsymbol{x}, \boldsymbol{x}') = \sum_{s'_j} P_{s_j, s'_j}(a_j, \hat{r}_j) V_\nu(\boldsymbol{x}').$$

To prove the desired results, we distinguish between the following cases.

- We first consider the case of $s_{1,j} = 0 < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j} = 0$. When $a_j = 1$ and for any $\boldsymbol{x}' - \{s'_j\}$, we have

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') = p_j V_\nu(\boldsymbol{x}'; s'_j = 1) + (1 - p_j) V_\nu(\boldsymbol{x}'; s'_j = 0),$$

$$U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') = \beta_j V_\nu(\boldsymbol{x}'; s'_j = s_{2,j} + 1) + (1 - \beta_j) V_\nu(\boldsymbol{x}'; s'_j = 0),$$

  where $V_\nu(\boldsymbol{x}'; s'_j = 0)$ is the estimated value function of the state $\boldsymbol{x}'$ with $s'_j = 0$ at iteration ν (at the risk of abusing the notation, we use $V(\boldsymbol{x}; s_j = s_1)$ and $V(\boldsymbol{x}; s_j = s_2)$ to represent the value functions of two states that differ only in the value of s_j). Then, we get

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') - U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') \leq (p_j - \beta_j)\big(V_\nu(\boldsymbol{x}'; s'_j = 1) - V_\nu(\boldsymbol{x}'; s'_j = 0)\big) \leq 0.$$

  The inequalities hold since $\beta_j > p_j$ and Lemma 1 are true at iteration ν by assumption. Therefore, we have $U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') \leq U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}')$ when $a_j = 1$ for any $\boldsymbol{x}' - \{s'_j\}$.
  For the case of $a_i = 1$ where $i \neq j$, we notice that $a_j = 0$. Then, for any $\boldsymbol{x}' - \{s'_j\}$, we obtain

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') = p_j V_\nu(\boldsymbol{x}'; s'_j = 1) + (1 - p_j) V_\nu(\boldsymbol{x}'; s'_j = 0),$$

$$U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') = (1 - p_j) V_\nu(\boldsymbol{x}'; s'_j = s_{2,j} + 1) + p_j V_\nu(\boldsymbol{x}'; s'_j = 0).$$

  Therefore, when $a_i = 1$, we have

$$U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') - U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}') \leq (2p_j - 1)\big(V_\nu(\boldsymbol{x}'; s'_j = 1) - V_\nu(\boldsymbol{x}'; s'_j = 0)\big) \leq 0.$$

  The inequalities hold since $2p_j - 1 < 0$ and Lemma 1 is true at iteration ν by assumption. Combining with the case of $a_j = 1$, $U^j_\nu(\boldsymbol{x}_1, \boldsymbol{x}') \leq U^j_\nu(\boldsymbol{x}_2, \boldsymbol{x}')$ holds for any

$\boldsymbol{x}' - \{s_j'\}$ under any feasible action. Since $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ differ only in the value of s_j and $C(\boldsymbol{x})$ is non-decreasing in s_i for $1 \leq i \leq N$, we can see that $V_{\nu+1}^{\boldsymbol{a}}(\boldsymbol{x}_1) \leq V_{\nu+1}^{\boldsymbol{a}}(\boldsymbol{x}_2)$ for any feasible $\boldsymbol{a}$. Then, by (A1), we can conclude that the lemma holds at iteration $\nu+1$ when $s_{1,j} = 0 < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j} = 0$.

- When $s_{1,j} = 0 < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j} = 1$, by replacing the β_j's in the above case with α_j's, we can achieve the same result.
- When $0 < s_{1,j} < s_{2,j}$ and $\hat{r}_{1,j} = \hat{r}_{2,j}$, we notice that

$$P_{s_{1,j},s_{1,j}+1}(a_j, \hat{r}_{1,j}) = P_{s_{2,j},s_{2,j}+1}(a_j, \hat{r}_{2,j}),$$
$$P_{s_{1,j},0}(a_j, \hat{r}_{1,j}) = P_{s_{2,j},0}(a_j, \hat{r}_{2,j}).$$

Then, leveraging the monotonicity of $V_\nu(\boldsymbol{x})$ and $C(\boldsymbol{x})$, we can conclude with the same result.

Combining the three cases, we prove that the lemma also holds at iteration $\nu+1$ of VIA. Therefore, the lemma holds at any iteration ν by mathematical induction. Since the results hold for any $1 \leq j \leq N$ and VIA is guaranteed to converge to the value function when $\nu \to +\infty$, we can conclude our proof.

Appendix B. Proof of Lemma 2

We inherit the notations in the proof of Lemma 1. We still use mathematical induction to obtain the desired results. The base case $\nu = 0$ is true by initialization. We assume the lemma holds at iterative ν and examine whether it still holds at iteration $\nu+1$. In the case of $M = 1$, we rewrite (5) as

$$V_{\nu+1}(\boldsymbol{x}) = \min_{1 \leq j \leq N} V_{\nu+1}^j(\boldsymbol{x}), \tag{A2}$$

where

$$V_{\nu+1}^j(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}'} \left\{ \left(\prod_{i \neq j} P_{x_i, x_i'}^i(0) \right) P_{x_j, x_j'}^j(1) V_\nu(\boldsymbol{x}') \right\}, \tag{A3}$$

and $P_{x,x'}^i(a_i)$ is the probability that action a_i will lead to state x' when user i is at state x. To get the desired results, we distinguish between the following cases

- We first show that $V_{\nu+1}^j(\boldsymbol{x}) = V_{\nu+1}^k(\mathscr{P}(\boldsymbol{x}))$. According to (A3), we have

$$V_{\nu+1}^j(\boldsymbol{x}) = C(\boldsymbol{x}) - \theta + \sum_{\boldsymbol{x}'} \left\{ \left(\prod_{i \neq j,k} P_{x_i, x_i'}^i(0) \right) P_{x_k, x_k'}^k(0) P_{x_j, x_j'}^j(1) V_\nu(\boldsymbol{x}') \right\}.$$

$$V_{\nu+1}^k(\mathscr{P}(\boldsymbol{x})) = C(\mathscr{P}(\boldsymbol{x})) - \theta + \sum_{\mathscr{P}(\boldsymbol{x})'} \left(\prod_{i \neq j,k} P_{\mathscr{P}(\boldsymbol{x})_i, \mathscr{P}(\boldsymbol{x})_i'}^i(0) \right) P_{\mathscr{P}(\boldsymbol{x})_k, \mathscr{P}(\boldsymbol{x})_k'}^k(1) P_{\mathscr{P}(\boldsymbol{x})_j, \mathscr{P}(\boldsymbol{x})_j'}^j(0) V_\nu(\mathscr{P}(\boldsymbol{x})').$$

It is obvious that for any $\mathscr{P}(\boldsymbol{x})'$, there always exists $\mathscr{P}(\boldsymbol{x}'') = \mathscr{P}(\boldsymbol{x})'$. Then, we obtain

$$\begin{aligned} V_{\nu+1}^k(\mathscr{P}(\boldsymbol{x})) &= C(\mathscr{P}(\boldsymbol{x})) - \theta + \sum_{\mathscr{P}(\boldsymbol{x}'')} \left(\prod_{i \neq j,k} P_{x_i, x_i''}^i(0) \right) P_{x_j, \mathscr{P}(\boldsymbol{x}'')_k}^k(1) P_{x_k, \mathscr{P}(\boldsymbol{x}'')_j}^j(0) V_\nu(\mathscr{P}(\boldsymbol{x}'')) \\ &= C(\mathscr{P}(\boldsymbol{x})) - \theta + \sum_{\boldsymbol{x}''} \left(\prod_{i \neq j,k} P_{x_i, x_i''}^i(0) \right) P_{x_j, x_j''}^k(1) P_{x_k, x_k''}^j(0) V_\nu(\boldsymbol{x}'') \\ &= C(\mathscr{P}(\boldsymbol{x})) - \theta + \sum_{\boldsymbol{x}'} \left(\prod_{i \neq j,k} P_{x_i, x_i'}^i(0) \right) P_{x_j, x_j'}^k(1) P_{x_k, x_k'}^j(0) V_\nu(\boldsymbol{x}'). \end{aligned}$$

The second equality follows from the definition of $\mathscr{P}(\cdot)$, the property of summation, and the assumption at iteration ν. The last equality follows from the variable renaming. Then, by the definition of statistically identical, we have $P^k_{x_j,x'_j}(1) = P^j_{x_j,x'_j}(1)$, $P^j_{x_k,x'_k}(0) = P^k_{x_k,x'_k}(0)$, and $C(\boldsymbol{x}) = C(\mathscr{P}(\boldsymbol{x}))$. Therefore, we can conclude that $V^j_{\nu+1}(\boldsymbol{x}) = V^k_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$.

- Along the same lines, we can easily show that $V^k_{\nu+1}(\boldsymbol{x}) = V^j_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$ and $V^i_{\nu+1}(\boldsymbol{x}) = V^i_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$ for $i \neq j,k$.

Combining the above cases with (A2), we prove that $V_{\nu+1}(\boldsymbol{x}) = V_{\nu+1}(\mathscr{P}(\boldsymbol{x}))$. Then, by induction, we have $V_\nu(\boldsymbol{x}) = V_\nu(\mathscr{P}(\boldsymbol{x}))$ at any iteration ν. Since VIA is guaranteed to converge to the value function when $\nu \to +\infty$, we can conclude our proof.

Appendix C. Proof of Theorem 1

For arbitrary j and k

$$\delta^{j,k}(\boldsymbol{x}) = \sum_{\boldsymbol{x}'-\{x'_j,x'_k\}} \left\{ \left(\prod_{i\neq j,k} P_{x_i,x'_i}(0) \right) \sum_{\hat{r}'_j,\hat{r}'_k} P(\hat{r}'_j)P(\hat{r}'_k)R^{j,k}(\boldsymbol{x},\boldsymbol{x}') \right\}, \tag{A4}$$

where

$$R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = \sum_{s'_j,s'_k} \left[\left(P_{s_k,s'_k}(0,\hat{r}_k)P_{s_j,s'_j}(1,\hat{r}_j) - P_{s_k,s'_k}(1,\hat{r}_k)P_{s_j,s'_j}(0,\hat{r}_j) \right) V(\boldsymbol{x}') \right]. \tag{A5}$$

With this in mind, we will prove the properties one by one.

Property 1—$\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $\hat{r}_k = p^0_{e,k} = 0$. The equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$.

When $\hat{r}_k = p^0_{e,k} = 0$, transmitting the update from user k will necessarily fail. Therefore, $P_{s_k,s'_k}(0,0) = P_{s_k,s'_k}(1,0)$ for any s_k and s'_k. Then, we have

$$R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = \sum_{s'_k} P_{s_k,s'_k}(0,0) \sum_{s'_j} \left[\left(P_{s_j,s'_j}(1,\hat{r}_j) - P_{s_j,s'_j}(0,\hat{r}_j) \right) V(\boldsymbol{x}') \right].$$

To identify the sign of $R^{j,k}(\boldsymbol{x},\boldsymbol{x}')$, we distinguish between the following cases

- When $s_j = 0$, we can easily show that $R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = 0$ for any $\boldsymbol{x}' - \{s'_j,s'_k\}$ by noticing that the two possible actions with respect to user j (i.e., $a_j = 1$ and $a_j = 0$) are equivalent when $s_j = 0$. Since $\delta^{j,k}(\boldsymbol{x})$ is a linear combination of $R^{j,k}(\boldsymbol{x},\boldsymbol{x}')$'s with non-negative coefficients, we can conclude that $\delta^{j,k}(\boldsymbol{x}) = 0$ in this case.
- When $s_j > 0$ and $\hat{r}_j = 1$, for any $\boldsymbol{x}' - \{s'_j,s'_k\}$, we have

$$\begin{aligned} R^{j,k}(\boldsymbol{x},\boldsymbol{x}') &= \sum_{s'_k} P_{s_k,s'_k}(0,0)(\alpha_j + p_j - 1)\big(V(\boldsymbol{x}';s'_j = s_j + 1) - V(\boldsymbol{x}';s'_j = 0)\big) \\ &\leq 0. \end{aligned} \tag{A6}$$

The inequality holds because of Lemma 1 and the fact that $\alpha_j + p_j < 1$. We recall that $\delta^{j,k}(\boldsymbol{x})$ is a linear combination of $R^{j,k}(\boldsymbol{x},\boldsymbol{x}')$'s with non-negative coefficients. Then, we can conclude that $\delta^{j,k}(\boldsymbol{x}) \leq 0$ in this case.
- When $s_j > 0$ and $\hat{r}_j = 0$, by replacing the α_j in (A6) with β_j, we can get the same result. In this case, the equality holds when $\beta_j + p_j = 1$, or, equivalently, $p^0_{e,j} = 0$.

Combining the cases, we prove the first property.

Property 2—$\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $\hat{r}_j$ and is non-decreasing in $\hat{r}_k$ when $s_j, s_k > 0$. At the same time, $\delta^{j,k}(\boldsymbol{x})$ is independent of $\hat{r}_i$ for any $i \neq j, k$.

We first prove the monotonicity of $\delta^{j,k}(\boldsymbol{x})$ with respect to $\hat{r}_j$. To this end, we define $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ as two states that differ only in the value of $\hat{r}_j$. Without a loss of generality, we assume $\hat{r}_{1,j} = 1$ and $\hat{r}_{2,j} = 0$. Then, we investigate the sign of $\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2)$. We define $x_i \triangleq x_{1,i} = x_{2,i}$ for $i \neq j$. Then, according to (A4), $\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2)$ can be written as

$$\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2) = \sum_{\boldsymbol{x}' - \{x'_j, x'_k\}} \left\{ \left(\prod_{i \neq j,k} P_{x_i, x'_i}(0) \right) \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j) P(\hat{r}'_k) \left(R^{j,k}(\boldsymbol{x}_1, \boldsymbol{x}') - R^{j,k}(\boldsymbol{x}_2, \boldsymbol{x}') \right) \right\}.$$

Since $x_{1,k} = x_{2,k}$, we have $P_{s_{1,k}, s'_k}(a, \hat{r}_{1,k}) = P_{s_{2,k}, s'_k}(a, \hat{r}_{2,k})$ for any s'_k. We recall that the transition probability is independent of $\hat{r}$ when $a = 0$. Combining with the fact that $s_{1,j} = s_{2,j}$, we also have $P_{s_{1,j}, s'_j}(0, \hat{r}_{1,j}) = P_{s_{2,j}, s'_j}(0, \hat{r}_{2,j})$ for any s'_j. Combining together, we obtain

$$P_{s_{1,k}, s'_k}(1, \hat{r}_{1,k}) P_{s_{1,j}, s'_j}(0, \hat{r}_{1,j}) = P_{s_{2,k}, s'_k}(1, \hat{r}_{2,k}) P_{s_{2,j}, s'_j}(0, \hat{r}_{2,j}),$$

$$P_{s_{1,k}, s'_k}(0, \hat{r}_{1,k}) = P_{s_{2,k}, s'_k}(0, \hat{r}_{2,k}).$$

Leveraging the above two problems, we have

$$R^{j,k}(\boldsymbol{x}_1, \boldsymbol{x}') - R^{j,k}(\boldsymbol{x}_2, \boldsymbol{x}') = \sum_{s'_j, s'_k} \left[P_{s_k, s'_k}(0, \hat{r}_k) \left(P_{s_{1,j}, s'_j}(1, \hat{r}_{1,j}) - P_{s_{2,j}, s'_j}(1, \hat{r}_{2,j}) \right) V(\boldsymbol{x}') \right].$$

Consequently, we obtain

$$\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2) = \sum_{\boldsymbol{x}' - \{x'_j\}} \left\{ \prod_{i \neq j} P_{x_i, x'_i}(0) \left[\sum_{\hat{r}'_j} P(\hat{r}'_j) \sum_{s'_j} \left(P_{s_{1,j}, s'_j}(1,1) - P_{s_{2,j}, s'_j}(1,0) \right) V(\boldsymbol{x}') \right] \right\}.$$

In the following, we characterize the sign of

$$R_1 \triangleq \sum_{s'_j} \left(P_{s_{1,j}, s'_j}(1,1) - P_{s_{2,j}, s'_j}(1,0) \right) V(\boldsymbol{x}').$$

As $s_{1,j} = s_{2,j} > 0$, for any $\boldsymbol{x}' - \{s'_j\}$, we have

$$R_1 = \big((1 - \alpha_j) - (1 - \beta_j) \big) V(\boldsymbol{x}'; s'_j = 0) + (\alpha_j - \beta_j) V(\boldsymbol{x}'; s'_j = s_{1,j} + 1) \leq 0.$$

The inequality follows from Lemma 1 and the fact that $\beta_j > \alpha_j$. Since $\delta^{j,k}(\boldsymbol{x}_1) - \delta^{j,k}(\boldsymbol{x}_2)$ is a linear combination of R_1's with non-negative coefficients, we can conclude that $\delta^{j,k}(\boldsymbol{x}_1) \leq \delta^{j,k}(\boldsymbol{x}_2)$. Since $\hat{r}_{1,j} > \hat{r}_{2,j}$, we can see that $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in $\hat{r}_j$.

In a very similar way, we can show that $\delta^{j,k}(\boldsymbol{x})$ is non-decreasing in $\hat{r}_k$. We recall that $\hat{r}_i$ will not affect the system dynamic if $a_i = 0$. Consequently, we can conclude that $\delta^{j,k}(\boldsymbol{x})$ is independent of $\hat{r}_i$ for any $i \neq j, k$.

Combining together, we prove the second property.

Property 3—$\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $s_k = 0$. The equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$.

Since the probabilities are non-negative, it is sufficient to show that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ satisfies Property 3 for any $\boldsymbol{x}' - \{s'_j, s'_k\}$. More precisely, it is sufficient to show that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \leq 0$

for any $\boldsymbol{x}' - \{s'_j, s'_k\}$ when $s_k = 0$ and the equality holds when $s_j = 0$ or $\hat{r}_j = p^0_{e,j} = 0$. We recall that $P_{s_k,s'_k}(1,\hat{r}_k) = P_{s_k,s'_k}(0,\hat{r}_k)$ for any s'_k when $s_k = 0$. Hence, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, we have

$$R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = \sum_{s'_k}\left[P_{s_k,s'_k}(0,\hat{r}_k)\sum_{s'_j}\left(P_{s_j,s'_j}(1,\hat{r}_j) - P_{s_j,s'_j}(0,\hat{r}_j)\right)V(\boldsymbol{x}')\right].$$

Then, we investigate the following quantity for any $\boldsymbol{x}' - \{s'_j\}$

$$R_2 \triangleq \sum_{s'_j}\left(P_{s_j,s'_j}(1,\hat{r}_j) - P_{x_j,x'_j}(0,\hat{r}_j)\right)V(\boldsymbol{x}').$$

To this end, we distinguish between the following cases

- When $s_j = 0$, we have $P_{s_j,s'_j}(1,\hat{r}_j) = P_{s_j,s'_j}(0,\hat{r}_j)$ for any s'_j. Thus, we conclude that $R_2 = 0$ for any $\boldsymbol{x}' - \{s'_j\}$. Consequently, $R^{j,k}(\boldsymbol{x},\boldsymbol{x}') = 0$ for any $\boldsymbol{x}' - \{s'_j, s'_k\}$.
- When $s_j > 0$ and $\hat{r}_j = 1$, for any $\boldsymbol{x}' - \{s'_j\}$, we have

$$R_2 = (\alpha_j - 1 + p_j)V(\boldsymbol{x}'; s'_j = s_j + 1) + (1 - \alpha_j - p_j)V(\boldsymbol{x}'; s'_j = 0) \leq 0 \tag{A7}$$

  The inequality follows from Lemma 1 and the fact that $\alpha_j + p_j < 1$. Thus, $R^{j,k}(\boldsymbol{x},\boldsymbol{x}') \leq 0$ for any $\boldsymbol{x}' - \{s'_j, s'_k\}$.
- When $s_j > 0$ and $\hat{r}_j = 0$, by replacing the α_j in (A7) with β_j, we can get the same result. In this case, the equality holds when $\beta_j + p_j = 1$, or, equivalently, $p^0_{e,j} = 0$.

Combined together, we can conclude that Property 3 is true.

Property 4—$\delta^{j,k}(\boldsymbol{x})$ is non-increasing in s_j if $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ and is non-decreasing in s_k if $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$ when $s_j, s_k > 0$. We define $\Gamma_i^1 \triangleq \frac{\alpha_i}{1-p_i}$ and $\Gamma_i^0 \triangleq \frac{\beta_i}{1-p_i}$ for $1 \leq i \leq N$.

Such as we did in the proof of Property 3, it is sufficient to show that $R^{j,k}(\boldsymbol{x},\boldsymbol{x}')$ satisfies Property 4 for any $\boldsymbol{x}' - \{s'_j, s'_k\}$. We recall that $R^{j,k}(\boldsymbol{x},\boldsymbol{x}')$ depends on the values of $\hat{r}_j$ and $\hat{r}_k$. Therefore, we distinguish between the following cases

- In the case of $\hat{r}_j = \hat{r}_k = 1$ and $s_j, s_k > 0$, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, (A5) can be written as

$$\begin{aligned}
R^{j,k}(\boldsymbol{x},\boldsymbol{x}') &= \sum_{s'_j,s'_k}\left[\left(P_{s_k,s'_k}(0,1)P_{s_j,s'_j}(1,1) - P_{s_k,s'_k}(1,1)P_{s_j,s'_j}(0,1)\right)V(\boldsymbol{x}')\right] \\
&= \big(p_k\alpha_j - (1-p_j)(1-\alpha_k)\big)V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = 0) \\
&\quad + \big((1-p_k)(1-\alpha_j) - p_j\alpha_k\big)V(\boldsymbol{x}'; s'_j = 0; s'_k = s_k + 1) \\
&\quad + \big((1-p_k)\alpha_j - (1-p_j)\alpha_k\big)V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = s_k + 1) \\
&\quad + \big(p_k(1-\alpha_j) - p_j(1-\alpha_k)\big)V(\boldsymbol{x}'; s'_j = 0; s'_k = 0).
\end{aligned}$$

  As we can verify

$$p_k\alpha_j - (1-p_j)(1-\alpha_k) < \frac{1}{2}(p_k + p_j - 1) < 0,$$

$$(1-p_k)(1-\alpha_j) - p_j\alpha_k > \frac{1}{2}(1 - p_k - p_j) > 0.$$

  We define $\Gamma_i^1 \triangleq \frac{\alpha_i}{1-p_i}$ and $\Gamma_i^0 \triangleq \frac{\beta_i}{1-p_i}$ for $1 \leq i \leq N$. Then, we have

$$\Gamma_j^1 \lesseqgtr \Gamma_k^1 \Longrightarrow (1-p_k)\alpha_j - (1-p_j)\alpha_k \lesseqgtr 0.$$

Combining with Lemma 1, we can conclude that, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ is non-increasing in s_j if $\Gamma_j^1 \leq \Gamma_k^1$ and is non-decreasing in s_k if $\Gamma_j^1 \geq \Gamma_k^1$.

- In the case of $\hat{r}_j = \hat{r}_k = 0$ and $s_j, s_k > 0$, by replacing the α's in the above case with β's, we can conclude with the same result.
- In the case of $\hat{r}_j = 1$, $\hat{r}_k = 0$, and $s_j, s_k > 0$, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, (A5) can be written as

$$\begin{aligned} R^{j,k}(\boldsymbol{x}, \boldsymbol{x}') &= \sum_{s'_j, s'_k} \left[\left(P_{s_k, s'_k}(0,0) P_{s_j, s'_j}(1,1) - P_{s_k, s'_k}(1,0) P_{s_j, s'_j}(0,1) \right) V(\boldsymbol{x}') \right] \\ &= \left(p_k \alpha_j - (1-p_j)(1-\beta_k)\right) V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = 0) \\ &\quad + \left((1-p_k)(1-\alpha_j) - p_j \beta_k\right) V(\boldsymbol{x}'; s'_j = 0; s'_k = s_k + 1) \\ &\quad + \left((1-p_k)\alpha_j - (1-p_j)\beta_k\right) V(\boldsymbol{x}'; s'_j = s_j + 1; s'_k = s_k + 1) \\ &\quad + \left(p_k(1-\alpha_j) - p_j(1-\beta_k)\right) V(\boldsymbol{x}'; s'_j = 0; s'_k = 0). \end{aligned}$$

As we can verify

$$p_k \alpha_j - (1-p_j)(1-\beta_k) < p_k \left(p_j - \frac{1}{2} \right) < 0,$$

$$(1-p_k)(1-\alpha_j) - p_j \beta_k > (1-p_k) \left(\frac{1}{2} - p_j \right) > 0.$$

At the same time

$$\Gamma_j^1 \lesseqgtr \Gamma_k^0 \Longrightarrow (1-p_k)\alpha_j - (1-p_j)\beta_k \lesseqgtr 0.$$

Combined with Lemma 1, we can conclude that, for any $\boldsymbol{x}' - \{s'_j, s'_k\}$, $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ is non-increasing in s_j if $\Gamma_j^1 \leq \Gamma_k^0$ and is non-decreasing in s_k if $\Gamma_j^1 \geq \Gamma_k^0$.

- In the case of $\hat{r}_j = 0$, $\hat{r}_k = 1$, and $s_j, s_k > 0$, by swapping the α's and β's in the above case, we can conclude with the same result.

Combined together, we conclude that $R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$ satisfies Property 3 for any $\boldsymbol{x}' - \{s'_j, s'_k\}$. Consequently, $\delta^{j,k}(\boldsymbol{x})$ is non-increasing in s_j if $\Gamma_j^{\hat{r}_j} \leq \Gamma_k^{\hat{r}_k}$ and is non-decreasing in s_k if $\Gamma_j^{\hat{r}_j} \geq \Gamma_k^{\hat{r}_k}$ when $s_j, s_k > 0$.

Property 5—$\delta^{j,k}(\boldsymbol{x}) \leq 0$ if $s_j \geq s_k$, $\hat{r}_j \geq \hat{r}_k$, and users j and k are statistically identical.

According to Property 3, it is sufficient to consider the case where $s_j, s_k > 0$. We notice that the sign of $\delta^{j,k}(\boldsymbol{x})$ can be captured by the sign of the quantity $Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') \triangleq \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j) P(\hat{r}'_k) R^{j,k}(\boldsymbol{x}, \boldsymbol{x}')$. Thus, we divide our discussion into the following cases.

- We first consider the case of $s_j \geq s_k > 0$ and $\hat{r}_j = \hat{r}_k = 0$. Leveraging the definition of statistically identical, for any $\boldsymbol{x}' - \{x'_j, x'_k\}$, we have

$$\begin{aligned} Q^{j,k}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{\hat{r}'_j, \hat{r}'_k} P(\hat{r}'_j) P(\hat{r}'_k) \kappa_1 \Big(& V(\boldsymbol{x}'; x'_j = (0, \hat{r}'_j); x'_k = (s_k + 1, \hat{r}'_k)) - \\ & V(\boldsymbol{x}'; x'_j = (s_j + 1, \hat{r}'_j); x'_k = (0, \hat{r}'_k)) \Big), \end{aligned}$$

where $\kappa_1 = 1 - p_j - \beta_j \geq 0$. Then, by substituting the values of $P(\hat{r})$ and using Lemma 2, we obtain

$$\begin{aligned}Q^{j,k}(\boldsymbol{x},\boldsymbol{x}') =&\gamma_j\gamma_k\kappa_1 V(\boldsymbol{x}';x'_j=(s_k+1,1);x'_k=(0,1))-\\&\gamma_j\gamma_k\kappa_1 V(\boldsymbol{x}';x'_j=(s_j+1,1);x'_k=(0,1))+\\&(1-\gamma_j)(1-\gamma_k)\kappa_1 V(\boldsymbol{x}';x'_j=(s_k+1,0);x'_k=(0,0))-\\&(1-\gamma_j)(1-\gamma_k)\kappa_1 V(\boldsymbol{x}';x'_j=(s_j+1,0);x'_k=(0,0))+\\&\gamma_k(1-\gamma_j)\kappa_1 V(\boldsymbol{x}';x'_j=(s_k+1,1);x'_k=(0,0))-\\&\gamma_k(1-\gamma_j)\kappa_1 V(\boldsymbol{x}';x'_j=(s_j+1,0);x'_k=(0,1))+\\&\gamma_j(1-\gamma_k)\kappa_1 V(\boldsymbol{x}';x'_j=(s_k+1,0);x'_k=(0,1))-\\&\gamma_j(1-\gamma_k)\kappa_1 V(\boldsymbol{x}';x'_j=(s_j+1,1);x'_k=(0,0)).\end{aligned}$$

Since users j and k are statistically identical, we have $\gamma_j=\gamma_k$. Then, by Lemma 1, we have $Q^{j,k}(\boldsymbol{x},\boldsymbol{x}')\leq 0$ for any $\boldsymbol{x}'-\{x'_j,x'_k\}$. Since $\delta^{j,k}(\boldsymbol{x})$ is a linear combination of $Q^{j,k}(\boldsymbol{x},\boldsymbol{x}')$'s with non-negative coefficients, we can conclude that $\delta^{j,k}(\boldsymbol{x})\leq 0$.

- For the case of $s_j\geq s_k>0$ and $\hat{r}_j=\hat{r}_k=1$, by replacing β_j in κ_1 with α_j, we can conclude with the same result.
- Then, we consider the case of $s_j\geq s_k>0$, $\hat{r}_j=1$, and $\hat{r}_k=0$. We first notice that, for any $\boldsymbol{x}'-\{s'_j,s'_k\}$

$$\begin{aligned}R^{j,k}(\boldsymbol{x},\boldsymbol{x}') =&\big(p_k\alpha_j-(1-p_j)(1-\beta_k)\big)V(\boldsymbol{x}';s'_j=s_j+1;s'_k=0)+\\&\big((1-p_k)(1-\alpha_j)-p_j\beta_k\big)V(\boldsymbol{x}';s'_j=0;s'_k=s_k+1)+\\&\big((1-p_k)\alpha_j-(1-p_j)\beta_k\big)V(\boldsymbol{x}';s'_j=s_j+1;s'_k=s_k+1)+\\&\big(p_k(1-\alpha_j)-p_j(1-\beta_k)\big)V(\boldsymbol{x}';s'_j=0;s'_k=0).\end{aligned}$$

As users j and k are statistically identical, we have $p_j=p_k$ and $\alpha_j<\beta_k$. Leveraging Lemma 1, we have

$$\begin{aligned}R^{j,k}(\boldsymbol{x},\boldsymbol{x}')\leq(\alpha_j+p_j-1)\Big(&V(\boldsymbol{x}';s'_j=s_j+1;s'_k=0)-\\&V(\boldsymbol{x}';s'_j=0;s'_k=s_k+1)\Big).\end{aligned}$$

Then, for any $\boldsymbol{x}'-\{x'_j,x'_k\}$

$$\begin{aligned}Q^{j,k}(\boldsymbol{x},\boldsymbol{x}')\leq\sum_{\hat{r}'_j,\hat{r}'_k}P(\hat{r}'_j)P(\hat{r}'_k)\kappa_2\Big(&V(\boldsymbol{x}';x'_j=(0,\hat{r}'_j);x'_k=(s_k+1,\hat{r}'_k))-\\&V(\boldsymbol{x}';x'_j=(s_j+1,\hat{r}'_j);x'_k=(0,\hat{r}'_k))\Big),\end{aligned}$$

where $\kappa_2=1-p_j-\alpha_j>0$. Such as we did in the previous cases, we can leverage Lemmas 1 and 2 to conclude that $Q^{j,k}(\boldsymbol{x},\boldsymbol{x}')\leq 0$ for any $\boldsymbol{x}'-\{x'_j,x'_k\}$. Consequently, $\delta^{j,k}(\boldsymbol{x})\leq 0$ in this case. The details are omitted for the sake of space.

Combined together, we conclude the proof of Property 5.

Appendix D. Proof of Corollary 2

We follow the same steps as in the proof of Lemma 1. To prove the corollary, it is sufficient to show that $V(x_1)\leq V(x_2)$ when $s_1<s_2$ and $\hat{r}_1=\hat{r}_2$. We use mathematical induction to prove the monotonicity. First of all, the base case (i.e., $\nu=0$) is true by initialization. We assume the lemma holds at iteration ν. Then, we want to examine

whether it holds at iteration $\nu+1$. For the system with a single user, the update step reported in problem (5) can be simplified and rewritten as follows

$$V_{\nu+1}(x) = \min_{a\in\{0,1\}} V^a_{\nu+1}(x), \tag{A8}$$

where

$$V^a_{\nu+1}(x) = C(x,a) - \theta + \sum_{\hat{r}'} P(\hat{r}') \sum_{s'} P_{s,s'}(a,\hat{r}) V_\nu(x'),$$

and θ is the optimal value for $\mathcal{M}_1(\lambda,-1)$. To prove the desired results, we distinguish between the following cases

- We first consider the case of $s_1 = 0 < s_2$ and $\hat{r}_1 = \hat{r}_2 = 0$. When $a = 1$, we have

$$V^1_{\nu+1}(x_1) = C(x_1,1) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big(pV_\nu(1,\hat{r}') + (1-p)V_\nu(0,\hat{r}') \Big),$$

$$V^1_{\nu+1}(x_2) = C(x_2,1) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big(\beta V_\nu(s_2+1,\hat{r}') + (1-\beta)V_\nu(0,\hat{r}') \Big).$$

  Subtracting the two expressions yields

$$\begin{aligned} &V^1_{\nu+1}(x_1) - V^1_{\nu+1}(x_2) \\ &\quad \le C(x_1,1) - C(x_2,1) + \sum_{\hat{r}'} P(\hat{r}') \Big[(p-\beta)\big(V_\nu(1,\hat{r}') - V_\nu(0,\hat{r}')\big) \Big] \le 0. \end{aligned}$$

  The inequalities hold since $\beta > p$, $C(x,a)$ is non-decreasing in s, and Corollary 2 is true at iteration ν by assumption.
  For the case of $a = 0$, we obtain

$$V^0_{\nu+1}(x_1) = C(x_1,0) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big(pV_\nu(1,\hat{r}') + (1-p)V_\nu(0,\hat{r}') \Big),$$

$$V^0_{\nu+1}(x_2) = C(x_2,0) - \theta + \sum_{\hat{r}'} P(\hat{r}') \Big((1-p)V_\nu(s_2+1,\hat{r}') + pV_\nu(0,\hat{r}') \Big).$$

  Therefore, when $a = 0$, we have

$$\begin{aligned} &V^0_{\nu+1}(x_1) - V^0_{\nu+1}(x_2) \\ &\quad \le C(x_1,0) - C(x_2,0) + \sum_{\hat{r}'} P(\hat{r}') \Big[(2p-1)\big(V_\nu(1,\hat{r}') - V_\nu(0,\hat{r}')\big) \Big] \le 0. \end{aligned}$$

  The inequalities hold since $2p - 1 < 0$, $C(x,a)$ is non-decreasing in s, and Corollary 2 is true at iteration ν by assumption. Combined together, we can see that $V^a_{\nu+1}(x_1) \le V^a_{\nu+1}(x_2)$ for any feasible a. Then, by problem (A8), we can conclude that the lemma holds at iteration $\nu+1$ when $s_1 = 0 < s_2$ and $\hat{r}_1 = \hat{r}_2 = 0$.
- When $s_1 = 0 < s_2$ and $\hat{r}_1 = \hat{r}_2 = 1$, by replacing the β's in the above case with α's, we can achieve the same result.
- When $0 < s_1 < s_2$ and $\hat{r}_1 = \hat{r}_2$, we notice that $P_{s_1,s_1+1}(a,\hat{r}_1) = P_{s_2,s_2+1}(a,\hat{r}_2)$ and $P_{s_1,0}(a,\hat{r}_1) = P_{s_2,0}(a,\hat{r}_2)$. Then, leveraging the monotonicity of $V_\nu(x)$ and $C(x,a)$, we can conclude with the same result.

Combining the three cases, we prove that the lemma holds at iteration $\nu+1$ of VIA. Therefore, the lemma holds at any iteration ν by mathematical induction. Since VIA is guaranteed to converge to the value function when $\nu \to +\infty$, we can conclude our proof.

Appendix E. Proof of Proposition 1

We define $\Delta V(x) \triangleq V^1(x) - V^0(x)$ where $V^a(x)$ is the value function resulting from taking action a at state x. Then, $V^a(x)$ can be calculated as follows

$$V^a(x) = C(x,a) - \theta + \sum_{x' \in \mathcal{X}} P_{x,x'}(a) V(x'), \tag{A9}$$

where θ is the optimal value for $\mathcal{M}_1(\lambda, -1)$. Hence, the optimal action at state x can be fully characterized by the sign of $\Delta V(x)$. More precisely, the optimal action at state x is $a = 1$ if $\Delta V(x) < 0$, and $a = 0$ is optimal otherwise. To determine the sign of $\Delta V(x)$ for each state, we distinguish between the following cases

- We first consider the state $x = (0, \hat{r})$. Applying the results in Section 2.3 to problem (A9), we obtain

$$\begin{aligned} V^0(0,\hat{r}) = &-\theta + (1-\gamma)(1-p)V(0,0) + (1-\gamma)pV(1,0) + \\ &\gamma(1-p)V(0,1) + \gamma p V(1,1), \end{aligned}$$

$$V^1(0,\hat{r}) = \lambda + V^0(0,\hat{r}). \tag{A10}$$

  Therefore, $\Delta V(0,\hat{r}) = \lambda \geq 0$. Thus, the optimal action at state $(0,\hat{r})$ is $a = 0$.
- Then, we consider the state $x = (s,0)$ where $s > 0$. Applying the results in Section 2.3 to Equation (A9), we obtain

$$\begin{aligned} V^0(s,0) = &f(s) - \theta + (1-\gamma)pV(0,0) + (1-\gamma)(1-p)V(s+1,0) + \\ &\gamma p V(0,1) + \gamma(1-p)V(s+1,1), \end{aligned}$$

$$\begin{aligned} V^1(s,0) = &f(s) + \lambda - \theta + (1-\gamma)(1-\beta)V(0,0) + (1-\gamma)\beta V(s+1,0) + \\ &\gamma(1-\beta)V(0,1) + \gamma\beta V(s+1,1). \end{aligned}$$

  Then,

$$\Delta V(s,0) = \lambda + p_e^0(1-2p)\omega, \tag{A11}$$

  where $\omega = (1-\gamma)[V(0,0) - V(s+1,0)] + \gamma[V(0,1) - V(s+1,1)] \leq 0$.
- Finally, we consider the state $x = (s,1)$ where $s > 0$. Following the same trajectory, we have

$$\Delta V(s,1) = \lambda + (1 - p_e^1)(1-2p)\omega.$$

According to Corollary 2 and the fact that $p < 0.5$, we can see that $\Delta V(s,0)$ and $\Delta V(s,1)$ are both a constant λ plus a term that is non-increasing in s. As the time penalty function is unbounded, the value function must also be unbounded. Then, combining the three cases, we can conclude the following. For fixed $\hat{r}$, there always exists a threshold $n_{\hat{r}} > 0$ such that the optimal action at state $(s,\hat{r})$ where $s \geq n_{\hat{r}}$ is $a = 1$, otherwise $a = 0$ is optimal. Since $\hat{r} \in \{0,1\}$, the optimal policy can be fully captured by the pair (n_0, n_1).

In the following, we determine the relationship between n_0 and n_1. We have

$$\Delta V(s,1) - \Delta V(s,0) = (1 - p_e^1 - p_e^0)(1-2p)\omega \leq 0.$$

At the same time, for the threshold n_0, we know $\Delta V(n_0,0) < 0$. Then, we have $\Delta V(n_0,1) \leq \Delta V(n_0,0) < 0$. Combined with the fact that $\Delta V(s,\hat{r})$ is non-increasing in s, we can conclude that the ordering $n_0 \geq n_1$ is true.

Appendix F. Proof of Proposition 2

We notice that the dynamic of AoII under threshold policy can be fully captured by a Discrete-Time Markov Chain (DTMC). Then, the expected AoII $\bar{\Delta}_{\boldsymbol{n}}$ and the expected transmission rate $\bar{\rho}_{\boldsymbol{n}}$ under threshold policy $\boldsymbol{n} = (n_0, n_1)$ can be obtained from the stationary

distribution of the induced DTMC. Let the states of the induced DTMC be the values of s. We recall that $\hat{r}$ is an independent Bernoulli random variable with parameter γ. Combined with the results in Section 2.3, we can easily obtain the state transition probabilities of the induced DTMC, which are shown in Figure A1.

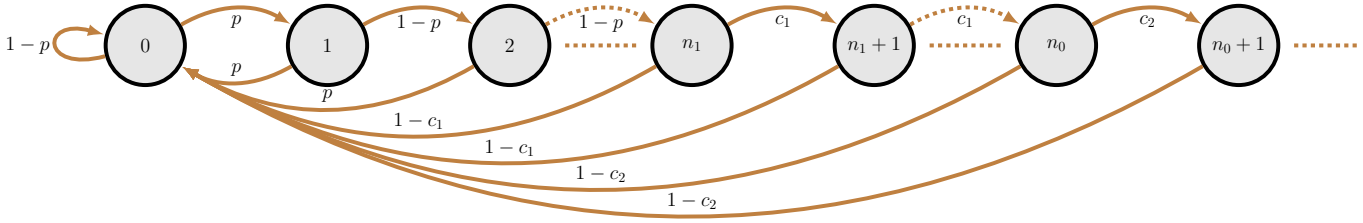


Figure A1. DTMC induced by the threshold policy $\boldsymbol{n} = (n_0, n_1)$. In the figure, $c_1 = (1-\gamma)(1-p) + \gamma\alpha$ and $c_2 = (1-\gamma)\beta + \gamma\alpha$.

The balance equations of the induced DTMC are the following

$$(1-p)\pi_0 + p\sum_{k=1}^{n_1-1}\pi_k + (1-c_1)\sum_{k=n_1}^{n_0-1}\pi_k + (1-c_2)\sum_{k=n_0}^{+\infty}\pi_k = \pi_0.$$

$$p\pi_0 = \pi_1.$$

$$(1-p)\pi_{k-1} = \pi_k \ \ for\ 2 \le k \le n_1.$$

$$c_1\pi_{k-1} = \pi_k \ \ for\ n_1+1 \le k \le n_0.$$

$$c_2\pi_{k-1} = \pi_k \ \ for\ n_0+1 \le k.$$

$$\sum_{k=0}^{+\infty}\pi_k = 1.$$

Then, we can easily solve the above system of linear equations. After some algebraic manipulation, we obtain the following

$$\pi_0 = \frac{1}{2 + p(1-p)^{n_1-1}\left[\frac{1}{1-c_1} - \frac{1}{p} + c_1^{n_0-n_1}\left(\frac{1}{1-c_2} - \frac{1}{1-c_1}\right)\right]}.$$

$$\pi_k = p(1-p)^{k-1}\pi_0 \ \ for\ 1 \le k \le n_1.$$

$$\pi_k = p(1-p)^{n_1-1}c_1^{k-n_1}\pi_0 \ \ for\ n_1+1 \le k \le n_0.$$

$$\pi_k = p(1-p)^{n_1-1}c_1^{n_0-n_1}c_2^{k-n_0}\pi_0 \ \ for\ n_0+1 \le k.$$

Equipped with the above results, we proceed with calculating $\bar{\Delta}_{\boldsymbol{n}}$ and $\bar{\rho}_{\boldsymbol{n}}$. According to problem (6a), the expected AoII is:

$$\bar{\Delta}_{\boldsymbol{n}} = \sum_{k=0}^{+\infty} f(k)\pi_k.$$

Substituting the expressions of π_k's, we can get the expression of $\bar{\Delta}_{\boldsymbol{n}}$. Proposition 1 tells us the following.

- For state $(s, \hat{r})$ where $s < n_1$, it is optimal to stay idle (i.e., $a = 0$).
- For state $(s, \hat{r})$ where $n_1 \le s < n_0$, it is optimal to make a transmission attempt only when $\hat{r} = 1$. We recall that $\hat{r}$ is an independent Bernoulli random variable with parameter γ. Therefore, the expected proportion of time that the system is at state $(s, 1)$ is $\gamma\pi_s$.

- For state $(s, \hat{r})$ where $s \geq n_0$, it is optimal to make transmission attempt regardless of $\hat{r}$.

Combined with problem (6b), we have

$$\bar{\rho}_{\boldsymbol{n}} = \gamma \sum_{k=n_1}^{n_0-1} \pi_k + \sum_{k=n_0}^{+\infty} \pi_k.$$

Substituting the expressions of π_k's, we can obtain the closed-form expression of $\bar{\rho}_{\boldsymbol{n}}$.

Appendix G. Proof of Proposition 4

We first tackle the Whittle's indexes at state $(0, \hat{r})$ and $(s, 0)$ where $s > 0$. To this end, we distinguish between the following cases

- We first consider the state $x = (0, \hat{r})$. By definition, Whittle's index is the infimum λ such that $V^0(x) = V^1(x)$. According to (A10), we can conclude that $W_x = 0$ when $x = (0, \hat{r})$.
- Then, we consider the state $x = (s, 0)$ where $s > 0$. We recall that $p_e^0 = 0$. Then, we can conclude, from (A11), that $W_x = 0$ for all $x = (s, 0)$ where $s > 0$.

Now, we tackle the Whittle's index at state $x = (s, 1)$ where $s > 0$. For convenience, we denote by W_n the Whittle's index at state $x = (n, 1)$. According to the monotonicity of $\Delta V(x)$ shown in the proof of Proposition 1, we can conclude that threshold policy $\boldsymbol{n} = (+\infty, n+1)$ is optimal when $V^0(n, 1) = V^1(n, 1)$. Then, we can prove the following

Lemma A1. *When (9) is satisfied and* $V^0(n, 1) = V^1(n, 1)$, $V(s, 1) = V(s, 0) \triangleq V(s)$ *for* $0 \leq s \leq n$.

Proof. Since the value function satisfies the Bellman equation, it is sufficient to show that $V(s, 1)$ and $V(s, 0)$ satisfy the same Bellman equation. We recall that the Bellman equation for $V(x)$ is given by

$$V(x) = \min_{a \in \{0,1\}} V^a(x),$$

where

$$V^a(x) = C(x, a) - \theta + \sum_{x'} P_{x,x'}(a) V(x'), \tag{A12}$$

and θ is the optimal value of the decoupled problem. We recall, from Corollary 3, that the optimal action at state $(s, 0)$ is staying idle (i.e., $a = 0$) for any s. We also know that threshold policy $\boldsymbol{n} = (+\infty, n+1)$ is optimal when $V^0(n, 1) = V^1(n, 1)$. Therefore, the optimal actions at states $(s, 0)$ and $(s, 1)$ where $s \leq n$ are the same (i.e., $a = 0$). Equivalently, we have

$$V(s, \hat{r}) = V^0(s, \hat{r}), \quad \textit{for } s \leq n. \tag{A13}$$

According to the system dynamic reported in Section 2.3, we know that the state transition probabilities are independent of $\hat{r}$ when $a = 0$. Meanwhile, $\hat{r}$ does not affect the instant cost. Let $x_1 = (s, 1)$ and $x_2 = (s, 0)$. Then, for any x', we have

$$P_{x_1,x'}(0) = P_{x_2,x'}(0).$$

$$C(x_1, 0) = C(x_2, 0).$$

Hence, according to (A12), we can see that $V^0(s, 0) = V^0(s, 1)$ for any $s \leq n$. Combined with problem (A13), we can conclude that $V(s, 0) = V(s, 1)$ for any $0 \leq s \leq n$. □

By definition, Whittle's index W_n is the infimum λ such that $V^0(n, 1) = V^1(n, 1)$. In this case, according to Lemma A1, $V(0, 1) = V(0, 0) = V(0)$. Then, $V^0(n, 1)$ and $V^1(n, 1)$ can be written as

$$V^0(n, 1) = f(n) - \theta + pV(0) + (1 - p)[(1 - \gamma)V(n + 1, 0) + \gamma V(n + 1, 1)]. \tag{A14}$$

$$V^1(n,1) = f(n) + W_n - \theta + (1-\alpha)V(0) + \alpha[(1-\gamma)V(n+1,0) + \gamma V(n+1,1)].$$

Without a loss of generality, we assume $V(0) = 0$. Then, equating the two expressions yields

$$W_n = (1-p-\alpha)(\gamma V(n+1,1) + (1-\gamma)V(n+1,0)). \tag{A15}$$

Combining problems (A14) and (A15), we conclude that W_n is

$$W_n = \frac{(1-p-\alpha)(V^0(n,1) + \theta - f(n))}{1-p}.$$

Since the optimal action at state $(n,1)$ is $a = 0$, we have $V^0(n,1) = V(n,1) = V(n)$. Finally, we obtain

$$W_n = \frac{(1-p-\alpha)(V(n) + \theta - f(n))}{1-p}. \tag{A16}$$

Now, we tackle the expression of $V(n)$. When $V^0(n,1) = V^1(n,1)$, the optimal action at state $(s,\hat{r})$ where $0 \le s < n$ is staying idle. Then, leveraging Lemma A1, value function $V(s)$ where $0 \le s < n$ satisfies the following

$$V(s) = \begin{cases} -\theta + f(0) + pV(1) & \text{when } s = 0, \\ -\theta + f(s) + (1-p)V(s+1) & \text{when } 0 < s < n. \end{cases} \tag{A17}$$

By backward induction, we end up with the following equation for $0 < s < n$.

$$V(s) = \frac{-\theta(1-(1-p)^{n-s})}{p} + \sum_{k=1}^{n-s} f(n-k)(1-p)^{n-s-k} + (1-p)^{n-s}V(n).$$

Letting $s = 1$ yields

$$V(1) = \frac{-\theta(1-(1-p)^{n-1})}{p} + \sum_{k=1}^{n-1} f(n-k)(1-p)^{n-1-k} + (1-p)^{n-1}V(n).$$

From problem (A17), $V(1)$ also satisfies the following

$$V(1) = \frac{\theta - f(0)}{p}.$$

Equating the two expressions of $V(1)$, we obtain

$$V(n) = \frac{-f(0)}{p(1-p)^{n-1}} + \theta\left(\frac{2}{p(1-p)^{n-1}} - \frac{1}{p}\right) - \sum_{k=1}^{n-1} f(n-k)(1-p)^{-k}. \tag{A18}$$

We recall that, when $V^0(n,1) = V^1(n,1)$, threshold policy $\boldsymbol{n} = (+\infty, n+1)$ is optimal and both actions at state $x = (n,1)$ are equally desirable. Thus, threshold policy $\boldsymbol{n} = (+\infty, n)$ is also optimal. Then, we know

$$\theta = \bar{\Delta}_n + W_n\bar{\rho}_n, \tag{A19}$$

where $\bar{\Delta}_n$ and $\bar{\rho}_n$ are the expected AoII and the expected transmission rate under threshold policy $\boldsymbol{n} = (+\infty, n)$, respectively. Finally, combining problems (A16), (A18) and (A19), we obtain

$$W_n = \frac{\dfrac{-f(0)}{p(1-p)^n} + \bar{\Delta}_n\dfrac{2-(1-p)^n}{p(1-p)^n} - (1-p)^{-n}\left(\displaystyle\sum_{k=1}^{n} f(k)(1-p)^{k-1}\right)}{\dfrac{1}{1-p-\alpha} - \bar{\rho}_n\dfrac{2-(1-p)^n}{p(1-p)^n}}.$$

After some algebraic manipulation, we have

$$W_n = \frac{(1-c_1)\sum_{k=n+1}^{+\infty} f(k)c_1^{k-n-1} - \bar{\Delta}_n}{\frac{(1-c_1)(1-p)-\gamma(1-p-\alpha)}{c_1(1-p-\alpha)} + \bar{\rho}_n},$$

where $c_1 = (1-\gamma)(1-p) + \gamma\alpha$.

In the following, we investigate some properties of Whittle's index. First of all, W_n is non-negative since $1-p-\alpha$ and $V(n+1,\hat{r})$ in (A15) are all non-negative. Meanwhile, combining (A15) with the fact that $V(n,\hat{r})$ is non-decreasing in n, we can verify that W_n is non-decreasing in n. Combined with the Whittle's indexes in two other cases (i.e., $x = (0,\hat{r})$ and $x = (s,0)$ where $s > 0$), we can easily obtain the properties of W_x as detailed in Proposition 4.

Appendix H. Proof of Proposition 5

We notice that $\mathcal{M}_1(\lambda,-1)$ coincides with the decoupled model studied in Section 4.2. When problem (9) is satisfied, the decoupled problem is indexable, and, according to Corollary 3, we only need to show that n is the optimal threshold for the states with $\hat{r} = 1$. We first tackle the case of $\lambda > 0$. To this end, we divide our discussion into the following cases

- For state $(s,1)$ where $s < n$, $W_s \leq \lambda$ by definition. As the problem is indexable, we have $D(W_s) \subseteq D(\lambda)$. We recall that $W_s \triangleq \min\{\lambda' \geq 0 : V^0(s,1) = V^1(s,1)\}$. Equivalently, $W_s \triangleq \min\{\lambda' \geq 0 : (s,1) \in D(\lambda')\}$. Then, we know $(s,1) \in D(W_s)$. Combined together, we conclude that $(s,1) \in D(\lambda)$. In other words, the optimal action at state $(s,1)$ where $s < n$ is to stay idle (i.e., $a = 0$).
- For state $(s,1)$ where $s \geq n$, we first recall that $W_s = \min\{\lambda' \geq 0 : (s,1) \in D(\lambda')\}$. Consequently, for any $\lambda' < W_s$, we know $(s,1) \notin D(\lambda')$. Meanwhile, we have $W_s \geq W_n > \lambda$ by the monotonicity of Whittle's index and the definition of n. Hence, we can conclude that $(s,1) \notin D(\lambda)$. In other words, the optimal action at state $(s,1)$ where $s \geq n$ is to make the transmission attempt.

Then, we conclude that n is the optimal threshold for the states with $\hat{r} = 1$ when $\lambda > 0$. In the case of $\lambda = 0$, according to the proof of Proposition 1, we can easily verify that the optimal threshold is 1.

Appendix I. Proof of Theorem 2

We first make the following definitions. When $\mathcal{M}_1(\lambda,-1)$ is at state x and action a is taken, cost $C_1(x,a) \triangleq f(s)$ and $C_2(x,a) \triangleq \lambda a$ are incurred. We denote the expected C_1-cost and the expected C_2-cost under policy ϕ as $\bar{C}_1(\phi)$ and $\bar{C}_2(\phi)$, respectively. Let G be a non-empty set of states. For the given state i, we define $\mathcal{R}^*(i,G)$ as the class of policies ϕ, for which the following hold

- The probability $P_\phi(x_n \in G \text{ for some } n \geq 1 \mid x_0 = i) = 1$ where x_n is the state of $\mathcal{M}_1(\lambda,-1)$ at time n.
- The expected time $m_{iG}(\phi)$ of a first passage from i to G under ϕ is finite.
- The expected C_1-cost $\bar{C}_1^{i,G}(\phi)$ and the expected C_2-cost $\bar{C}_2^{i,G}(\phi)$ of a first passage form i to G under ϕ are finite.

With the definitions in mind, we proceed with verifying the assumptions given in [27].

1. *For all $d > 0$, the set $A(d) = \{x \mid$ there exists an action a such that $C_1(x,a) + C_2(x,a) \leq d\}$ is finite*: For any state x, the cost satisfies $C_1(x,a) + C_2(x,a) = f(s) + \lambda a \geq f(s)$. The equality holds when $a = 0$. Then, the states in $A(d)$ must satisfy $f(s) \leq d$. Combined with the fact that $f(s)$ is a non-decreasing and unbounded function when $s \in \mathbb{N}_0$, we can conclude that $A(d)$ is finite.
2. *There exists a stationary policy e such that the induced Markov chain has the following properties: the state space $\mathcal{S}$ consists of a single (non-empty) positive recurrent class R and a*

set U of transient states such that $e \in \mathcal{R}^*(i,R)$ for $i \in U$. Moreover, both $\bar{C}_1(e)$ and $\bar{C}_2(e)$ on R are finite: We consider the policy under which the base station makes a transmission attempt at every time slot. According to the system dynamic detailed in Section 2.3, we can see that all the states communicate with state $(0,0)$ and $(0,0)$ communicates with all other states. Thus, the state space $\mathcal{S}$ consists of a single (non-empty) positive recurrent class and the set of transient states can simply be an empty set. $\bar{C}_1(e)$ and $\bar{C}_2(e)$ are trivially finite as we can verify using Proposition 2.

3. *Given any two state $x \neq y$, there exists a policy ϕ such that $\phi \in \mathcal{R}^*(x,y)$*: We notice that, under any policy, the maximum increase of s between two consecutive time slots is 1. Meanwhile, when s decreases, it decreases to zero. Combined with the fact that $\hat{r}$ is an independent Bernoulli random variable, we can conclude that there always exists a path between any x and y with positive probability. $m_{xy}(\phi)$, $\bar{C}_1^{x,y}(\phi)$, and $\bar{C}_2^{x,y}(\phi)$ are trivially finite.
4. *If a stationary policy ϕ has at least one positive recurrent state, then it has a single positive recurrent class R. Moreover, if $x = (0,0) \notin R$, then $\phi \in \mathcal{R}^*(x,R)$*: Given that $\hat{r}$ is an independent Bernoulli random variable, we can easily conclude from the system dynamic that all the states communicate with state $(0,0)$ and $(0,0)$ communicates with all other states under any stationary policy. Therefore, any positive recurrent class must contain state $(0,0)$. Thus, there must have only one positive recurrent class which is $R = \mathcal{S}$.
5. *There exists a policy ϕ such that $\bar{C}_1(\phi) < \infty$ and $\bar{C}_2(\phi) < K$ where $K \in (0,1]$*: We notice that $\bar{C}_1(\phi)$ and $\bar{C}_2(\phi)$ are nothing but the expected AoII and the expected transmission rate achieved by ϕ, respectively. Then, we can easily verify that such policy exists using Proposition 2.

As the assumptions are verified, we proceed with introducing the optimal randomized policy for given λ. We say a policy is λ-optimal if the policy is optimal for $\mathcal{M}_1(\lambda,-1)$. We consider two monotone sequences $\lambda_+^n \downarrow \lambda$ and $\lambda_-^n \uparrow \lambda$. Then, there exist subsequences of λ_+^n and λ_-^n such that the corresponding sequences of optimal policies converge. Then, according to Lemma 3.7 of [27], the limit points, denoted by $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$, are both λ-optimal. By Proposition 3.2 of [27], the Markov chains induced by $\boldsymbol{n}_{\lambda_+}$ and $\boldsymbol{n}_{\lambda_-}$ both contain a single non-empty positive recurrent class and state $(0,0)$ is positive recurrent in both induced Markov chains. Hence, the base station can choose which policy to follow each time the system reaches state $(0,0)$ while keeping the resulting randomized policy λ-optimal as suggested by Lemma 3.9 of [27]. More precisely, we consider the following randomized policy: each time the system reaches state $(0,0)$, the base station will choose $\boldsymbol{n}_{\lambda_-}$ with probability μ and $\boldsymbol{n}_{\lambda_+}$ with probability $1-\mu$. The chosen policy will be followed until the next choice. We denote such policy as $\boldsymbol{n}_\lambda$ and conclude that $\boldsymbol{n}_\lambda$ is λ-optimal under any $\mu \in [0,1]$.

Appendix J. Proof of Proposition 6

The value function $V(\boldsymbol{x})$ and $V^i(x_i)$ must satisfy their own Bellman equations. More precisely

$$V(\boldsymbol{x}) + \theta = \min_{\boldsymbol{a} \in \mathcal{A}_N(-1)} \left\{ C(\boldsymbol{x},\boldsymbol{a}) + \sum_{\boldsymbol{x}'} Pr(\boldsymbol{x}' \mid \boldsymbol{x},\boldsymbol{a}) V(\boldsymbol{x}') \right\},$$
$$V^i(x_i) + \theta_i = \min_{a_i \in \{0,1\}} \left\{ C(x_i,a_i) + \sum_{x_i'} Pr(x_i' \mid x_i,a_i) V^i(x_i') \right\}, \tag{A20}$$

where θ and θ_i are the optimal values of $\mathcal{M}_N(\lambda,-1)$ and $\mathcal{M}_1^i(\lambda,-1)$, respectively. We recall from Section 2.3 that the users are independent when action $\boldsymbol{a}$ and current state $\boldsymbol{x}$ are given. Thus

$$Pr(\boldsymbol{x}' \mid \boldsymbol{x},\boldsymbol{a}) = \prod_{i=1}^{N} Pr(x_i' \mid \boldsymbol{x},\boldsymbol{a}),$$

where $\boldsymbol{x}' = (x_1', \ldots, x_N')$. Then, we have

$$\sum_{\boldsymbol{x}'-\{x_i'\}} Pr(\boldsymbol{x}'-\{x_i'\} \mid \boldsymbol{x},\boldsymbol{a}) = \sum_{\boldsymbol{x}'-\{x_i'\}} \prod_{j\neq i} Pr(x_j' \mid \boldsymbol{x},\boldsymbol{a}) = 1.$$

We also recall from Section 2.3 that the state of user i depends only on its previous state and the action with respect to user i. Thus

$$Pr(x_i' \mid \boldsymbol{x},\boldsymbol{a}) = Pr(x_i' \mid x_i, a_i).$$

Combined together, we obtain

$$\begin{aligned}\sum_{i=1}^{N}\sum_{x_i'} Pr(x_i' \mid x_i,a_i)V^i(x_i') &= \sum_{i=1}^{N}\sum_{x_i'}\left[\sum_{\boldsymbol{x}'-\{x_i'\}}\prod_{j\neq i} Pr(x_j' \mid \boldsymbol{x},\boldsymbol{a})\right] Pr(x_i' \mid x_i,a_i)V^i(x_i') \\ &= \sum_{i=1}^{N}\sum_{x_i'}\left(\sum_{\boldsymbol{x}'-\{x_i'\}}\prod_{i=1}^{N} Pr(x_i' \mid \boldsymbol{x},\boldsymbol{a})V^i(x_i')\right) \\ &= \sum_{\boldsymbol{x}'} Pr(\boldsymbol{x}' \mid \boldsymbol{x},\boldsymbol{a})\left(\sum_{i=1}^{N} V^i(x_i')\right).\end{aligned} \tag{A21}$$

Then, we sum problem (A20) over all users which yields

$$\sum_{i=1}^{N}(V^i(x_i)+\theta_i) = \min_{\boldsymbol{a}}\left\{\sum_{i=1}^{N}\left(C(x_i,a_i)+\sum_{x_i'} Pr(x_i' \mid x_i,a_i)V^i(x_i')\right)\right\}.$$

We recall that $C(\boldsymbol{x},\boldsymbol{a}) = \sum_{i=1}^{N} C(x_i,a_i)$ by definition. Then, leveraging problem (A21), we obtain

$$\sum_{i=1}^{N} V^i(x_i) + \sum_{i=1}^{N}\theta_i = \min_{\boldsymbol{a}\in\mathcal{A}_N(-1)}\left\{C(\boldsymbol{x},\boldsymbol{a})+\sum_{\boldsymbol{x}'} Pr(\boldsymbol{x}' \mid \boldsymbol{x},\boldsymbol{a})\left(\sum_{i=1}^{N} V^i(x_i')\right)\right\}.$$

Since the solution to the Bellman equation is unique [21], we must have $\sum_{i=1}^{N} V^i(x_i) = V(\boldsymbol{x})$ and $\sum_{i=1}^{N}\theta_i = \theta$. Then, we can conclude that it is optimal for $\mathcal{M}_N(\lambda,-1)$ if each user adopts its own optimal policy.

Appendix K. Proof of Theorem 3

In this proof, we class a policy as λ^*-optimal if it is optimal for $\mathcal{M}_N(\lambda^*,-1)$. In Section 4.2, we ensure that, for each user, there exists at least one threshold policy that yields a finite expected AoII. Therefore, we can conclude that, for RP, there exists at least one policy that causes the expected AoII and the expected transmission rate to be both finite. Then, according to Lemma 3.10 of [27], a policy is optimal for RP if

1. It is λ^*-optimal;
2. The resulting expected transmission rate is equal to M.

We first construct a policy ϕ_{λ^*} that is λ^*-optimal. We recall from Proposition 6 that a policy is λ^*-optimal if it consists of the optimal policies for each $\mathcal{M}_1^i(\lambda^*,-1)$ where $1 \leq i \leq N$. According to Theorem 2, for any i, there exist $\boldsymbol{n}_{\lambda_-^*,i}$ and $\boldsymbol{n}_{\lambda_+^*,i}$ that are both optimal for $\mathcal{M}_1^i(\lambda^*,-1)$. Then, we can construct the policy ϕ_{λ^*} in the following way.

- For user i with $\boldsymbol{n}_{\lambda_-^*,i} = \boldsymbol{n}_{\lambda_+^*,i} \triangleq \boldsymbol{n}_{\lambda^*,i}$, the threshold policy $\boldsymbol{n}_{\lambda^*,i}$ is used. Then, the deterministic policy $\boldsymbol{n}_{\lambda^*,i}$ is optimal for $\mathcal{M}_1^i(\lambda^*,-1)$ and

  $$\bar{\rho}^i(\lambda^*) = \bar{\rho}^i(\lambda_-^*) = \bar{\rho}^i(\lambda_+^*).$$

  In this case, the choice of μ_i makes no difference.
- For user i with $\boldsymbol{n}_{\lambda_-^*,i} \neq \boldsymbol{n}_{\lambda_+^*,i}$, the randomized policy $\boldsymbol{n}_{\lambda^*,i}$ as detailed in Theorem 2 is used. Then, for any $\mu_i \in [0,1]$, the randomized policy $\boldsymbol{n}_{\lambda^*,i}$ is optimal for $\mathcal{M}_1^i(\lambda^*,-1)$ and

$$\bar{\rho}^i(\lambda^*) = \mu_i \bar{\rho}^i(\lambda^*_-) + (1-\mu_i)\bar{\rho}^i(\lambda^*_+).$$

Combing the two cases, we conclude that $\phi_{\lambda^*} = [\boldsymbol{n}_{\lambda^*,1}, \ldots, \boldsymbol{n}_{\lambda^*,N}]$ is λ^*-optimal under any $\mu_i \in [0,1]$. Hence, as long as the chosen μ_i's realize $\sum_{i=1}^N \bar{\rho}^i(\lambda^*) = M$, we can conclude that the randomized policy ϕ_{λ^*} is optimal for RP.

Appendix L. Proof of Proposition 8

We notice that $\mathcal{M}_1^i(\lambda^*, -1)$ coincides with the decoupled model studied in Section 4.2. Therefore, we can use the results in Section 4.2 to prove the properties. Since the users share the same structure, we ignore the user index i for simplicity. According to the definition of I_x, we have

$$\begin{aligned} I_x &= \sum_{x'} P_{x,x'}(0)V(x') - \sum_{x'} P_{x,x'}(1)V(x') - \lambda^* \\ &= -\Delta V(x). \end{aligned}$$

Leveraging the results in the proof of Proposition 1, we have the following

- For state $x = (0, \hat{r})$, $I_x = -\lambda^*$.
- For state $x = (s, 0)$ where $s > 0$, $I_x = -\lambda^* - p_e^0(1-2p)\omega$ where $\omega = (1-\gamma)[V(0,0) - V(s+1,0)] + \gamma[V(0,1) - V(s+1,1)] \leq 0$.
- For state $x = (s, 1)$ where $s > 0$, $I_x = -\lambda^* - (1-p_e^1)(1-2p)\omega$.

From the above three cases, we can easily conclude that $I_x \geq -\lambda^*$ and the equality holds when $\hat{r} = p_e^0 = 0$ or $s = 0$. As is proven in Corollary 2, $V(x)$ is non-decreasing in s. Hence, we can conclude that I_x is also non-decreasing in s. To show that I_x is monotone in $\hat{r}$, we consider two states $x_1 = (s, 1)$ and $x_2 = (s, 0)$. Then, we have

$$I_{x_2} - I_{x_1} = \Delta V(s,1) - \Delta V(s,0) = (1 - p_e^1 - p_e^0)(1-2p)\omega \leq 0.$$

Therefore, we can conclude that I_x is non-decreasing in $\hat{r}$.

Appendix M

Algorithm A1 Improved Relative Value Iteration

Require:
MDP $\mathcal{M} = (\mathcal{X}, \mathcal{P}, \mathcal{A}, \mathcal{C})$
Convergence Criteria ϵ

procedure RELATIVEVALUEITERATION($\mathcal{M}$,ϵ)
  Initialize $V_0(x) = 0$; $\nu = 0$
  Choose $x^{ref} \in \mathcal{X}$ arbitrarily
  while V_ν is not converged (RVI converges when the maximum difference between the results of two consecutive iterations is less than ϵ) **do**
    for $x = (s, \hat{r}) \in \mathcal{X}$ **do**
      if $\exists$ active state $(s_1, \hat{r}_1)$ s.t. $s_1 \leq s$ and $\hat{r}_1 \leq \hat{r}$ **then**
        $a^*(x) = 1$
        $Q_{\nu+1}(x) = C(x,1) + \sum_{x'} P_{xx'}(1)V_\nu(x')$
      else
        for $a \in \mathcal{A}$ **do**
          $H_{x,a} = C(x,a) + \sum_{x'} P_{xx'}(a)V_\nu(x')$
        $a^*(x) = \arg\min_a \{H_{x,a}\}$
        $Q_{\nu+1}(x) = H_{x,a^*}$
    for $x \in \mathcal{X}$ **do**
      $V_{\nu+1}(x) = Q_{\nu+1}(x) - Q_{\nu+1}(x^{ref})$
    $\nu = \nu + 1$
  return $\boldsymbol{n} \leftarrow a^*(x)$

Algorithm A2 Bisection Search

Require:
Maximum updates per transmission attempt M
MDP $\mathcal{M}_N(\lambda, -1) = (\mathcal{X}_N, \mathcal{A}_N(-1), \mathcal{P}_N, \mathcal{C}_N(\lambda))$
Tolerance ξ
Convergence criteria ϵ

procedure BISECTIONSEARCH($\mathcal{M}_N(\lambda, -1)$, M, ξ, ϵ)
  Initialize $\lambda_- = 0$; $\lambda_+ = 1$
  $\phi_{\lambda_+} \leftarrow (\mathcal{M}_N(\lambda_+, -1), \epsilon)$ using Section 5.1 and Proposition 6
  $\bar{\rho}(\lambda_+) \leftarrow \phi_{\lambda_+}$ using Proposition 2
  while $\bar{\rho}(\lambda_+) \geq M$ **do**
    $\lambda_- = \lambda_+$; $\lambda_+ = 2\lambda_+$
    $\phi_{\lambda_+} \leftarrow (\mathcal{M}_N(\lambda_+, -1), \epsilon)$ using Section 5.1 and Proposition 6
    $\bar{\rho}(\lambda_+) \leftarrow \phi_{\lambda_+}$ using Proposition 2
  while $\lambda_+ - \lambda_- \geq 2\xi$ **do**
    $\lambda = \frac{\lambda_+ + \lambda_-}{2}$
    $\phi_\lambda \leftarrow (\mathcal{M}_N(\lambda, -1), \epsilon)$ using Section 5.1 and Proposition 6
    $\bar{\rho}(\lambda) \leftarrow \phi_\lambda$ using Proposition 2
    if $\bar{\rho}(\lambda) > M$ **then**
      $\lambda_- = \lambda$
    else
      $\lambda_+ = \lambda$
  return $(\lambda_+^*, \lambda_-^*) \leftarrow (\lambda_+, \lambda_-)$

References

1. Maatouk, A.; Kriouile, S.; Assaad, M.; Ephremides, A. The age of incorrect information: A new performance metric for status updates. *IEEE/ACM Trans. Netw.* **2020**, *28*, 2215–2228. [CrossRef]
2. Uysal, E.; Kaya, O.; Ephremides, A.; Gross, J.; Codreanu, M.; Popovski, P.; Assaad, M.; Liva, G.; Munari, A.; Soleymani, T.; et al. Semantic communications in networked systems. *arXiv* **2021**, arXiv:2103.05391.
3. Kam, C.; Kompella, S.; Ephremides, A. Age of incorrect information for remote estimation of a binary markov source. In Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Toronto, ON, Canada, 6–9 July 2020; pp. 1–6.
4. Maatouk, A.; Assaad, M.; Ephremides, A. The age of incorrect information: An enabler of semantics-empowered communication. *arXiv* **2020**, arXiv:2012.13214.
5. Chen, Y.; Ephremides, A. Minimizing Age of Incorrect Information for Unreliable Channel with Power Constraint. *arXiv* **2021**, arXiv:2101.08908.
6. Kriouile, S.; Assaad, M. Minimizing the Age of Incorrect Information for Real-time Tracking of Markov Remote Sources. *arXiv* **2021**, arXiv:2102.03245.
7. Kadota, I.; Sinha, A.; Uysal-Biyikoglu, E.; Singh, R.; Modiano, E. Scheduling policies for minimizing age of information in broadcast wireless networks. *IEEE/ACM Trans. Netw.* **2018**, *26*, 2637–2650. [CrossRef]
8. Hsu, Y.P. Age of information: Whittle index for scheduling stochastic arrivals. In Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT), Vail, CO, USA, 17–22 June 2018; pp. 2634–2638.
9. Tripathi, V.; Modiano, E. A whittle index approach to minimizing functions of age of information. In Proceedings of the 2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 24–27 September 2019; pp. 1160–1167.
10. Maatouk, A.; Kriouile, S.; Assad, M.; Ephremides, A. On the optimality of the Whittle's index policy for minimizing the age of information. *IEEE Trans. Wirel. Commun.* **2020**, *20*, 1263–1277. [CrossRef]
11. Sun, J.; Jiang, Z.; Krishnamachari, B.; Zhou, S.; Niu, Z. Closed-form Whittle's index-enabled random access for timely status update. *IEEE Trans. Commun.* **2019**, *68*, 1538–1551. [CrossRef]
12. Nguyen, G.D.; Kompella, S.; Kam, C.; Wieselthier, J.E. Information freshness over a Markov channel: The effect of channel state information. *Ad Hoc Networks* **2019**, *86*, 63–71. [CrossRef]
13. Talak, R.; Karaman, S.; Modiano, E. Optimizing age of information in wireless networks with perfect channel state information. In Proceedings of the 2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Shanghai, China, 7–11 May 2018; pp. 1–8.
14. Shi, L.; Cheng, P.; Chen, J. Optimal periodic sensor scheduling with limited resources. *IEEE Trans. Autom. Control* **2011**, *56*, 2190–2195. [CrossRef]
15. Leong, A.S.; Dey, S.; Quevedo, D.E. Sensor scheduling in variance based event triggered estimation with packet drops. *IEEE Trans. Autom. Control* **2016**, *62*, 1880–1895. [CrossRef]

16. Mo, Y.; Garone, E.; Casavola, A.; Sinopoli, B. Stochastic sensor scheduling for energy constrained estimation in multi-hop wireless sensor networks. *IEEE Trans. Autom. Control* **2011**, *56*, 2489–2495. [CrossRef]
17. Kaul, S.; Yates, R.; Gruteser, M. Real-time status: How often should one update? In Proceedings of the 2012 Proceedings IEEE INFOCOM, Orlando, FL, USA, 25–30 March 2012; pp. 2731–2735.
18. Leong, A.S.; Ramaswamy, A.; Quevedo, D.E.; Karl, H.; Shi, L. Deep reinforcement learning for wireless sensor scheduling in cyber–physical systems. *Automatica* **2020**, *113*, 108759. [CrossRef]
19. Wang, J.; Ren, X.; Mo, Y.; Shi, L. Whittle index policy for dynamic multichannel allocation in remote state estimation. *IEEE Trans. Autom. Control* **2019**, *65*, 591–603. [CrossRef]
20. Gittins, J.; Glazebrook, K.; Weber, R. *Multi-Armed Bandit Allocation Indices*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
21. Russell, S.; Norvig, P. *Artificial Intelligence: A Modern Approach*, 3rd ed.; Prentice Hall Press: Hoboken, NJ, USA, 2009.
22. Whittle, P. Restless bandits: Activity allocation in a changing world. *J. Appl. Probab.* **1988**, *25*, 287–298. [CrossRef]
23. Weber, R.R.; Weiss, G. On an index policy for restless bandits. *J. Appl. Probab.* **1990**, *27*, 637–648. [CrossRef]
24. Glazebrook, K.D.; Ruiz-Hernandez, D.; Kirkbride, C. Some indexable families of restless bandit problems. *Adv. Appl. Probab.* **2006**, *38*, 643–672. [CrossRef]
25. Larrañaga, M. Dynamic Control of Stochastic and Fluid Resource-Sharing Systems. Ph.D. Thesis, Université de Toulouse, Toulouse, France, 2015.
26. Sennott, L.I. On computing average cost optimal policies with application to routing to parallel queues. *Math. Methods Oper. Res.* **1997**, *45*, 45–62. [CrossRef]
27. Sennott, L.I. Constrained average cost Markov decision chains. *Probab. Eng. Inf. Sci.* **1993**, *7*, 69–83. [CrossRef]
28. Bertsimas, D.; Niño-Mora, J. Restless bandits, linear programming relaxations, and a primal-dual index heuristic. *Oper. Res.* **2000**, *48*, 80–90. [CrossRef]
29. Littman, M.L.; Dean, T.L.; Kaelbling, L.P. On the complexity of solving Markov decision problems. *arXiv* **2013**, arXiv:1302.4971.
30. Verloop, I.M. Asymptotically optimal priority policies for indexable and nonindexable restless bandits. *Ann. Appl. Probab.* **2016**, *26*, 1947–1995. [CrossRef]