Efficient and Timely Memory Access

Vishakha Ramani, Ivan Seskar, Roy D. Yates
WINLAB, Rutgers University
Email: {vishakha, seskar, ryates}@winlab.rutgers.edu

***Abstract*—This paper investigates the optimization of memory sampling in status updating systems, where source updates are published in shared memory, and reader process samples the memory for source updates by paying a sampling cost. We formulate a discrete-time decision problem to find a sampling policy that minimizes average cost comprising age at the client and the cost incurred due to sampling. We establish that an optimal policy is a stationary and deterministic threshold-type policy, and subsequently derive optimal threshold and the corresponding optimal average cost.**

I. Introduction

This work examines status updating systems in which sources generate time-stamped status updates of a process of interest, and these updates are stored/written in a memory system. A reader fulfills clients' requests for these updates by reading from the memory. The asynchronous nature of reader-writer interactions within memory systems introduces significant challenges. In particular, the readers' memory accesses should be optimized for timely processing of source updates as the reader becomes aware of fresher updates in the memory only when it chooses to query the memory. Furthermore, the memory access process must be regulated by a synchronization method between readers and writers to avoid race conditions.

The primary question in this paper is when should the reader sample the memory. Typically, there is a cost associated with memory sampling, and this cost structure varies between systems. In systems with substantial object sizes, retrieving and locally copying objects incurs a high cost, while querying for timestamps remains relatively inexpensive. In contrast, there are systems where memory contains smaller objects, and the cost of retrieval is comparable to the cost of a timestamp query. These are systems where queries are sent to a distant database, with the cost being the latency associated with the query.

In this work, we focus on former class of systems where the Reader knows the freshness of object in the memory by virtue of inexpensive timestamp retrievals. However, due to longer read times, denoted by high sampling costs, the Reader must decide if sampling is justified compared to age reduction obtained after sampling.

A. Related Work

Prior research on timely memory access has explored issues related to the impact of different synchronization primitives on the timely retrieval of stored data items. Particularly, [1], [2] examined the impact of lock-based and lock-less synchronization primitives in the context of a timely packet forwarding application.

In [3], the authors addressed the timely processing of updates stored in memory from multiple sources. The framework involved a system where the Reader samples the memory as a renewal point process. Under this model, it was shown that a lazy sampling policy from the Reader proved to be an optimal strategy. The rationale behind this finding was that lazy sampling effectively mitigates the negative impacts of high variance in the client's processing times. This work differs from [3] in that it no longer assumes renewal sampling from the Reader and in contrast to previous work, in this study, the Reader pays a cost of sampling. Nevertheless, the primary objective remains consistent: to identify an optimal reading policy for the Reader that maximizes the timeliness of source updates with the client.

We note that the concept of timely memory sampling, wherein the Reader incurs a cost for sampling for age reduction, shares similarities with research focused on managing access for multiple users within a communication channel. Various studies in the AoI literature have explored Whittle's index-based transmission scheduling algorithms [4]–[13], wherein the scheduling problem is decomposed into multiple independent subproblems. Within each subproblem, an additional cost (C) is associated with updating the user. The goal is to determine, in each time slot, whether updating the user is warranted, thereby striking a balance between the updating cost and the age-related costs.

B. Contributions and Paper Outline

This paper investigates the relation between sampling costs and Age-of-Information [14]. In section II, we formulate our problem as a Markov Decision Process (MDP) with the goal of minimizing average cost comprising age at the client and the cost incurred due to sampling. In section III, we establish that an optimal policy of the MDP is a stationary and deterministic threshold-type policy. We then derive optimal threshold and the optimal average cost by exploiting the structure of optimal policy. Finally, section IV presents numerical evaluation on average cost against system parameters.

II. System Model

In this work, we focus on a class of systems (see Fig. 1) where a Writer writes the time-varying data received from the source into the memory, and a Reader samples the memory on behalf of a client. The client can be either the same entity as the Reader, running as a single process, or a separate process. Additionally, while the system will have many sources, our focus will be on the memory that tracks the status of a single

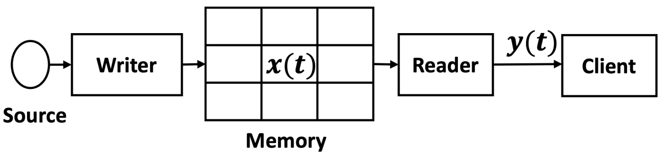

Fig. 1. A writer updates memory based on the update received from source. A client requests the Reader process to read the source updates from the memory. The source update publication in the memory generates age process $x(t)$, and the update sampling by the Reader generates age process $y(t)$ at the client input.

process of interest. Even in this seemingly straightforward scenario, not previously explored in the AoI literature, optimizing AoI presents non-trivial challenges.

We consider a discrete-time slotted system with slots labelled $t = 0, 1, 2, \ldots$. The system involves two key processes: writing the time-varying data from the source into the memory and reading the source data from memory. The modeling details of these processes are discussed below.

A. Writing source updates to memory

We assume the Writer commits/writes fresh (age zero) source updates to memory at the end of each slot with probability p, independent from slot to slot. These source updates generate the age process $x(t)$ in the memory.

In practice, the write time will be non-negligible. However, our focus in this work is not on systems where writing to the memory is the bottleneck process. Instead, our primary interest lies in examining the delays associated with reading and processing of source updates. Note that in the event that these writes do require time $\tau > 0$, $x(t)$ and the update age process at the client will be shifted by τ.

B. Sampling source updates from memory

At each time slot, the Reader determines whether to access the memory and read a source update. The update in memory is read over a period of a slot, and the reader gets the data at the end of the slot. Notably, this model aligns with the Read-Copy-Update (RCU) [15], [16] memory access paradigm, where a new update can be written in slot t while the Reader is in the process of reading the current update in the same slot.

The Reader generates an age process $y(t)$ at the input to the client that is a sampled version of source update age process $x(t)$ in the memory. Hence we say the Reader *samples* the updates in the memory.

The state-dependent action $a(t)$ selected by the Reader at time slot t determines whether the Reader remains idle ($a(t) = 0$) or performs a read operation ($a(t) = 1$). We consider a scenario where a non-negative fixed cost c is associated with reading the memory during each time slot. Ideally, the Reader aims to minimize $y(t)$, which means it would prefer to read in every slot to stay close to the age process $x(t)$. However, this comes at the cost of paying the sampling cost c. If the Reader samples too frequently, it might end up with the same update, resulting in no age reduction but incurring a penalty for sampling. On the contrary, if it reads too infrequently, the age at the client input increases.

In this work, we assume that the Reader is notified when an update is published in the memory, enabling the Reader to know the update age in the memory. Based on the system state, the Reader implements a scheduling scheme that minimizes the average cost $\mathrm{E}[y(t) + ca(t)]$. To address this, we model our problem as a Markov Decision Process (MDP).

C. Markov Decision Process Formulation

In the context of our MDP model, denoted with $\mathcal{M}$ from here on, the following four components make up the structure:

- States: We denote the set of possible system states by S which does not vary with time. State $s(t) \in S$ is a tuple $(x(t), y(t))$, where at the start of a time slot, $x(t) \in \{0, 1, 2, \ldots\}$ is the age of the update in the memory, and $y(t) \in \{1, 2, 3, \ldots\}$ is the age of sampled source updates at the client. Notice that S is a countably infinite set since age is unbounded.
- Action: Let $a(t) \in \{0, 1\}$ denote the action taken in slot t indicating Reader's decision, where $a(t) = 1$ if Reader decides to read and $a(t) = 0$ if idle.
- Transition Probabilities: Letting $\bar{p} = 1 - p$, when $a(t) = 1$, the transition probability from state $s = (x, y)$ to state $s' \in S$ is
$$\mathrm{P}[s' \mid s = (x, y), a = 1] = \begin{cases} p & s' = (0, x+1), \\ \bar{p} & s' = (x+1, x+1). \end{cases} \tag{1a}$$
And when $a(t) = 0$, the transition probability is
$$\mathrm{P}[s' \mid s = (x, y), a = 0] = \begin{cases} p & s' = (0, y+1), \\ \bar{p} & s' = (x+1, y+1). \end{cases} \tag{1b}$$
- Cost: The cost $C(s(t); a(t))$ incurred in state $s(t)$ in time slot t under action $a(t)$ is defined as:
$$C(s(t) = (x, y);\ a(t) = a) := y + ca. \tag{2}$$

Let $\pi = \{a(0), a(1), \ldots\}$ denote a policy that specifies an action $a(t)$ at slot t. The expected average cost under policy π starting from a given initial state at $t = 0$, $s(0) = s$, is defined as:
$$\begin{aligned} g_\pi(s) &= \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\left[\sum_{t=0}^{T-1} C(s(t); a(t)) \mid s(0) = s\right], \\ &= \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}_\pi\left[\sum_{t=0}^{T-1} (y(t) + ca(t))\right]. \end{aligned} \tag{3}$$

We say that policy π^* is average-cost optimal if $g_{\pi^*}(s) = \inf g_\pi(s)$ for every $s \in S$. We focus on the case where for some constant g, $g_{\pi^*}(s) = g$ for all $s \in S$. Thus, the problem is to obtain π^* such that $g = g_{\pi^*}(s) = \inf g_\pi(s)$ for every $s \in S$.

Our cost minimization problem falls within the category of average cost minimization problems. Given that the age can grow unbounded, both the number of states and the cost in each stage are countably infinite. In such Markov Decision

Processes (MDPs), the existence of an optimal policy, whether stationary or non-stationary, is not guaranteed [17, Chap 5]. Notably, even the existence of an optimal stationary policy may not hold, while an optimal non-stationary policy might exist [18].

Analyzing average cost problems with an infinite state space poses inherent difficulties. However, under certain conditions and structures, it is possible to develop useful results. Proving the existence of an optimal average cost stationary policy is not an immediate goal in this paper and we defer this discussion to later in Section V. There, we draw upon results from [19], which provides conditions ensuring the existence of an expected average cost optimal stationary policy. We verify that these conditions hold for our problem. In the subsequent section, we derive results regarding the structure of the optimal policy under the assumption that the optimal policy exists and the relative cost Bellman's equation is valid.

III. Characterization of cost optimality

A. Discounted Cost

We begin by introducing the α-discounted version of the problem. Recall that the state for MDP $\mathcal{M}$ is a tuple $s = (x, y)$, and $a \in \{0, 1\}$. Then using (1), the discounted cost Bellman's optimality equation for $\mathcal{M}$ is given by

$$\begin{aligned} V(x,y) = \min\{&y + \alpha\left(pV(0,y+1) + \bar{p}V(x+1,y+1)\right), \\ &y + c + \alpha(pV(0,x+1) + \bar{p}V(x+1,x+1))\}. \quad (4)\end{aligned}$$

Here, the first term of $\min$ corresponds to the reader staying idle ($a = 0$), and the second term corresponds to the reader sampling ($a = 1$). The action that is a minimizer of (4) is referred to as the *α-optimal action* and the resulting policy π^*_α is referred to as the *α-optimal policy*.

We define the value iteration $V_n(s)$ by $V_0(s) = 0, \forall s \in S$, and, for any $n > 0$,

$$\begin{aligned} V_{n+1}(x,y) = \min\{&y + \alpha\left(pV_n(0,y+1) + \bar{p}V_n(x+1,y+1)\right), \\ &y + c + \alpha(pV_n(0,x+1) + \bar{p}V_n(x+1,x+1))\}, \quad (5)\end{aligned}$$

For non-negative costs, it is evident that $V_n(s) \leq V_{n+1}(s)$. It then follows from [17, Theorem 4.2, Chapter III] that

$$\lim_{n\to\infty} V_n(s) = V(s), \qquad s \in S. \quad (6)$$

We now state properties of the value function.

Proposition 1. *(Monotonicity): The value function $V(x, y)$ is non-decreasing in both x and y.*

The proof, using mathematical induction on (5), is straightforward but omitted because of space constraints.

Proposition 2. *If the α-optimal action is to sample in (x, y), then the α-optimal action is to sample in every (x, y') with $y' \geq y$.*

Proof of this proposition is provided in the Appendix. Another version of this proposition asserts that if the α-optimal action is to sample in state (x, y) at stage n, then it is also optimal to sample in every (x, y') with $y' \geq y$ at stage n. The proof employing the value iteration (5) is omitted as it is similar to that of Proposition 2.

Proposition 3. *(Concavity): For a fixed x, $V(x, y + 1) - V(x, y)$ is non-increasing in y.*

The proof appears in the Appendix. The intuitive structure of the optimal policy is that with knowledge of the age in the memory, the Reader should refrain from sampling if the reduction in age doesn't justify the sampling cost. To further characterize this intuition, we introduce the following proposition. The proof appears in the Appendix.

Proposition 4. *If the α-optimal action in state (x, y) is to idle, then the α-optimal action in states $(x + i, y + i), \forall i \geq 1$ is to stay idle.*

Specifically, when the memory is freshly updated, the Reader must assess whether sampling is worthwhile. If it opts against sampling initially, it should consistently abstain from sampling in subsequent slots until the memory undergoes another update, as the age reduction remains constant in the absence of changes. In terms of the MDP $\mathcal{M}$, this concept translates to making a decision in the state $(0, y)$. If the optimal decision is not to sample at this point, then the Reader should consistently refrain from sampling in states $(1, y + 1)$, $(2, y + 2)$, and so on.

B. Average Cost Optimality

Since the conditions of Theorem 2 (in section V) hold, the cost-optimal policy π^* is the limit point of α-optimal policies π^*_α with $\alpha \to 1$ [19, Lemma]. Therefore, Propositions 2 and 4 are sufficient to provide the structure of average cost optimal policy. Specifically, Propositions 2 and 4 imply that there exists a threshold Y_0 such that it is optimal to sample in $(0, y)$ for every $y \geq Y_0$ and idle otherwise.

At this point, it is important to mention the set of feasible states under π^*. With $Y_0 = 1$, the optimal policy dictates sampling in every state $(0, y)$ where $y \geq 1$. Upon sampling in $(0, 1)$, the system transitions to feasible states, specifically $\{(0, 1), (1, 1)\}$. In state $(1, 1)$, a close examination of Bellman's equation (4) reveals that it is optimal to idle. Therefore, the set of possible states when choosing to idle in $(1, 1)$ becomes $\{(0, 2), (2, 2)\}$. Subsequent transitions follow a pattern where sampling in $(0, y)$ leads to states $\{(0, 1), (1, 1)\}$, and choosing to idle in states (i, i) with $i \in \mathbb{N}$ resulting in $\{(0, i + 1), (i + 1, i + 1)\}$.

In scenarios where $Y_0 > 1$, optimality dictates idling in $(0, y)$ with $y < Y_0$, prompting the system to transition to states $\{(0, y + 1), (1, y + 1)\}$. The subsequent action in $(0, y + 1)$ hinges upon whether $y + 1 < Y_0$. If $y + 1 \geq Y_0$, the system resets, transitioning to either $(0, 1)$ or $(1, 1)$; conversely, if $y + 1 < Y_0$, the system perpetuates a structure akin to that observed in state $(0, y)$. Conversely, if the system transitions to $(1, y + 1)$, idling in $(1, y + 1)$ is optimal. The resulting permissible states from this point include $\{(0, y + 2), (2, y + 2)\}$, and this pattern repeats. We summarize this set of feasible states for the optimal policy in the following proposition.

Proposition 5. *For MDP $\mathcal{M}$, under the optimal policy π^* with threshold Y_0, the set of feasible states is*

$$S^* = \{(0, y) \mid y \in \mathbb{N}\} \cup \{(x, y) \mid x \geq 1 \text{ and } y - x < Y_0\}. \quad (7)$$

To determine the optimal threshold for an optimal policy π^*, we employ the relative cost Bellman's equation

$$\begin{aligned} g + f(x, y) = \min\{&y + pf(0, y+1) + \bar{p}f(x+1, y+1), \\ &y + c + pf(0, x+1) + \bar{p}f(x+1, x+1)\}. \quad (8) \end{aligned}$$

Here, g denotes the optimal average cost, and $f(x, y)$ represents the relative cost-to-go function. Our objective is to identify relative cost-to-go function $f(x, y)$ for $(x, y) \in S^*$, facilitating the determination of the optimal threshold and, consequently, the optimal average cost.

Proposition 6. *Defining $(0, 1)$ as the reference state with $f(0, 1) = 0$, the relative cost functions satisify:*

(i) $f(0, Y_0 + 1) - f(0, Y_0) = 1$.

(ii) *For any $x \geq 0$,*

$$f(x, Y_0 - 1) = \frac{1}{p}(J_0 + \frac{\bar{p}}{p}) - 1, \quad (9)$$

where $J_0 = Y_0 - g + pf(0, Y_0)$.

(iii) *For every $y < Y_0$, $f(0, y) = f(1, y) \ldots = f(y, y)$.*

(iv) *When $Y_0 > 1$, $f(0, Y_0) = Y_0 - g + c$.*

The proof appears in the Appendix. We now use Proposition 6 to derive the optimal threshold.

Lemma 1. *As a function of the threshold Y_0, the average cost is*

$$g_0(Y_0) = \frac{1}{2}\left(\frac{1}{p} + Y_0 + \frac{2cp + \bar{p}/p}{pY_0 + \bar{p}}\right). \quad (10)$$

The proof appears in the Appendix.

Theorem 1. *The optimal threshold Y_0^* associated with optimal policy π^* for MDP $\mathcal{M}$ is $Y_0^* = \lceil Y' \rceil$ where*

$$Y' = \sqrt{2c + (1/p - 1/2)^2} - (1/p - 1/2). \quad (11)$$

Proof. It follows from (10) and some algebra that

$$g_0(Y_0) - g_0(Y_0 + 1) = \frac{-p^2}{2}\left[\frac{Y_0^2 + (2/p - 1)Y_0 - 2c}{(pY_0 + \bar{p})(pY_0 + 1)}\right] \quad (12)$$

We define $Q(Y_0) \equiv Y_0^2 + Y_0\,(2/p - 1) - 2c$ and we observe that Y' in (11) is the only positive root of $Q(y)$. Further $Q(Y_0) > 0$ for $Y_0 > Y'$. It then follows from (12) that $g_0(\lfloor Y' \rfloor) \geq g_0(\lceil Y' \rceil)$ and that $g_0(\lceil Y' \rceil), g_0(\lceil Y' \rceil + 1), \ldots$ is a non-decreasing sequence. □

Lemma 2. *The optimal average cost satisfies*

$$g \geq 1/2 + \sqrt{2c + 1/p^2} - 1/p. \quad (13)$$

Proof. Note that $g = \min_{Y_0 \in \mathbb{N}} g_0(Y_0) \geq \min_{y \in \mathbb{R}^+} g_0(y)$. To minimize $g_0(y)$ over positive reals, we set $dg_0(y)/dy = 0$, yielding $y = \tilde{Y}_0^* = -\bar{p}/p + \sqrt{2c + \bar{p}/p^2}$. This yields $g \geq g_0(\tilde{Y}_0^*)$, which is the lower bound (13). □

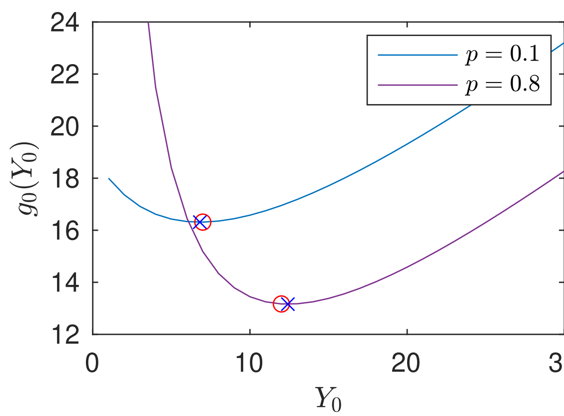

Fig. 2. Plot of average cost $g_0(Y_0)$ as a function of threshold Y_0 with sampling cost $c = 80$. The circle point is the minimum of $g_0(Y_0)$, the cross is the approximate optimal threshold $\tilde{Y}_0^*$.

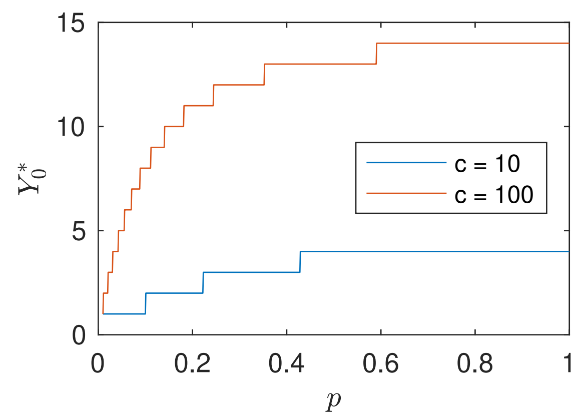

Fig. 3. Plot of optimal threshold Y_0^* as a function of probability p of source update publication in a slot, with a fixed sampling cost c.

IV. Numerical Evaluation

Figure 2 illustrates the behavior of the average cost given by (10) with respect to Y_0. For fixed system parameters p and c, the plot reveals an initial decline in average cost as threshold Y_0 increases. This trend aligns with expectations, as a low threshold prompts the Reader to sample too frequently, incurring the sampling cost without achieving a significant reduction in age, ultimately leading to a higher average cost. As Y_0 increases further, the cost of sampling approaches the gain obtained with age reduction. However, setting threshold Y_0 too high precludes timely access to the memory, resulting in increased age at the client and consequently increased average cost of the system. The observed behavior in Figure 2 highlights the existence of an optimal threshold at which the cost payed for sampling is justified with the corresponding reduction in age.

Fig. 3 illustrates the value of optimal threshold Y_0^* as a function of probability p of source update publication in a slot. We observe that the optimal threshold increases with p. When the Reader is required to make a decision in a given slot, it assesses both the age at the client and the age in the memory. These evaluations contribute to determining the potential age

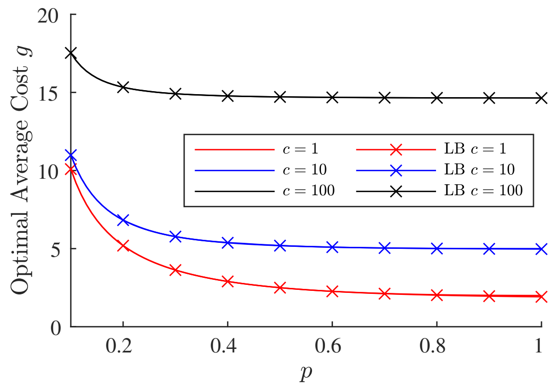

Fig. 4. Comparison of Optimal average cost g and the corresponding lower bound (LB) as a function of probability p of source update publication in a slot, with a fixed sampling cost c.

reduction vs the cost of sampling. In scenarios where the client's update is deemed sufficiently recent, the Reader may choose to skip sampling. This decision is influenced by a higher probability (p) of obtaining a more recent update soon, that will perhaps be worth sampling.

Figure 4 compares the optimal average cost g with the lower bound provided in (13). The tightness of the lower bound is evident, as it closely aligns with the curve of the optimal average cost. Additionally, the figure illustrates that the optimal cost tends to increase with an increase in the sampling cost c. This suggests that while designing the cost structure, the cost should be sufficiently high but not excessively so. Furthermore, the plot shows that the average cost decreases as the probability p of memory updates in a slot increases. This is intuitive, as frequent memory updates increase the likelihood of the Reader receiving a fresh update when it samples, thereby reducing the age at the client.

V. Stationary average cost optimal policy

In this section we verify that the average cost optimality equation for MDP holds for $\mathcal{M}$. To get started, we need the following result.

Lemma 3. *Under the deterministic stationary policy θ of reading in every slot, the system exhibits an irreducible, ergodic Markov Chain, with expected cost $M(x,y)$ of first passage from state $s=(x,y)$ to $(0,1)$ satisfying*

$$M(x,y) \leq \frac{1+p}{p^2}(c+y) + \frac{3}{2p^3}. \tag{14}$$

Proof of Lemma 3 appears in the Appendix. We now employ the lemma in verifying the conditions of the following theorem.

Theorem 2. *[19, Theorem] If the following conditions hold for MDP $\mathcal{M}$:*

1) *For every state s and discount factor α, the quantity $V(s)$ is finite,*
2) *$f_\alpha(s) := V(s) - V(0)$ satisfies $-N \overset{(a)}{\leq} f_\alpha(s) \overset{(b)}{\leq} M(s)$, where $M(s) \geq 0$, and*
3) *For all s and a, $\sum_{s'} \mathbb{P}_{s,s'}(a)M(s') < \infty$,*

then there exists a stationary policy that is average cost optimal for MDP $\mathcal{M}$. Moreover, for $\mathcal{M}$, there exists a constant $g = \lim_{\alpha\to 1}(1-\alpha)V(s)$ for every state s, and a function $f(s)$ with $-N \leq f(s) \leq M(s)$ that solve relative-cost Bellman's equation,

$$g + f(s) = \min_a \{C(s;a) + \sum_{s'\in S} \mathbb{P}_{s,s'}(a) f(s')\}. \tag{15}$$

For MDP $\mathcal{M}$, we choose reference state 0 as $(0,1)$. A sufficient condition for 1 and $2(b)$ to hold is the existence of a single stationary policy that induces an irreducible, ergodic Markov Chain, with the associated expected cost of first passage from any state (x,y) to state $(0,1)$ being finite ([19, Propositions 4 and 5]). Lemma 3 verifies that this sufficient condition is met for our problem. A sufficient condition for $2(a)$ is that $V(s)$ is non-decreasing in s [19]. Proposition 1 demonstrates that this sufficient condition is also met.

Now, condition 3 of Theorem 2 asserts that under any a, the quantity $\sum_{s'} \mathbb{P}_{s,s'}(a)M(s')$ should be finite. For MDP $\mathcal{M}$, from (1b), when $a=0$, we have for any state $s=(x,y)$,

$$\sum_{s'} \mathbb{P}_{s,s'}(0)M(s') = pM(0,y+1) + \bar{p}M(x+1,y+1). \tag{16}$$

From (1a), when $a=1$, we similarly have for any state $s=(x,y)$,

$$\sum_{s'} \mathbb{P}_{s,s'}(1)M(s') = pM(0,x+1) + \bar{p}M(x+1,x+1). \tag{17}$$

It follows from (16), (17) and Lemma 3 that condition 3 holds for MDP $\mathcal{M}$. Therefore, there exists a constant $g = \lim_{\alpha\to 1}(1-\alpha)V(x,y)$ for every state (x,y) that is an optimal average cost and a relative cost to go function $f(x,y)$ with $0 \leq f(x,y) \leq M(x,y)$.

VI. Conclusion

This paper focused on a class of systems where source updates are disseminated using shared memory. The Writer process records these source updates in the memory, and a Reader fulfills clients' requests for these measurements by reading from the memory. We studied the problem of optimizing memory access by the Reader with respect to minimizing average cost. Our main contributions included establishing the existence of an optimal stationary deterministic policy for our Markov Decision Process (MDP). Furthermore, we demonstrated that the optimal policy has a threshold structure.

A key insight from our analysis was that the Reader should choose to sample only when the memory undergoes an update. If the Reader decides not to sample, this decision of staying idle should perpetuate in subsequent slots until the memory is updated with a fresh source update. This is because, in the absence of updates, there is no change in age reduction; it remains the same as when the memory was last updated. Finally, an interesting extension to this work would involve investigating an alternate regime for optimizing memory access, where the Reader operates without knowledge of source update age in the memory.

REFERENCES

[1] V. Ramani, J. Chen, and R. D. Yates, "Lock-based or lock-less: Which is fresh?" 2023.

[2] ——, "Timely mobile routing: An experimental study," 2023.

[3] V. Ramani, I. Seskar, and R. D. Yates, "Timely processing of updates from multiple sources," 2023.

[4] Y.-P. Hsu, "Age of information: Whittle index for scheduling stochastic arrivals," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 2634–2638.

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

[6] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling policies for minimizing age of information in broadcast wireless networks," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2637–2650, 2018.

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

[8] J. Sun, Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Closed-form whittle's index-enabled random access for timely status update," *IEEE Transactions on Communications*, vol. 68, no. 3, pp. 1538–1551, 2020.

[9] I. Kadota, A. Sinha, and E. Modiano, "Scheduling algorithms for optimizing age of information in wireless networks with throughput constraints," *IEEE/ACM Transactions on Networking*, vol. 27, no. 4, pp. 1359–1372, 2019.

[10] V. Tripathi and E. Modiano, "A whittle index approach to minimizing functions of age of information," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 1160–1167.

[11] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, "Asymptotically optimal scheduling policy for minimizing the age of information," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 1747–1752.

[12] J. Sun, Z. Jiang, S. Zhou, and Z. Niu, "Optimizing information freshness in broadcast network with unreliable links and random arrivals: An approximate index policy," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. IEEE, 2019, pp. 115–120.

[13] Z. Jiang, B. Krishnamachari, S. Zhou, and Z. Niu, "Can decentralized status update achieve universally near-optimal age-of-information in wireless multiaccess channels?" in *2018 30th International Teletraffic Congress (ITC 30)*, vol. 1. IEEE, 2018, pp. 144–152.

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

[15] P. E. McKenney and J. D. Slingwine, "Read-copy update: Using execution history to solve concurrency problems," in *Parallel and Distributed Computing and Systems*, vol. 509518. Citeseer, 1998, pp. 509–518.

[16] P. E. Mckenney, J. Appavoo, A. Kleen, O. Krieger, O. Krieger, R. Russell, D. Sarma, and M. Soni, "Read-copy update," in *In Ottawa Linux Symposium*, 2001, pp. 338–367.

[17] S. Ross, *Introduction to Stochastic Dynamic Programming*, ser. Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press, 1983. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978012598420150009X

[18] ——, *Applied Probability Models with Optimization Applications*, ser. Holden-Day series in industrial engineering and management science. Holden-Day, 1970.

[19] L. I. Sennott, "Average cost optimal stationary policies in infinite state markov decision processes with unbounded costs," *Operations Research*, vol. 37, no. 4, pp. 626–633, 1989. [Online]. Available: http://www.jstor.org/stable/171262

APPENDIX

PROOF OF LEMMA 3

Lemma 3. The expected first passage cost $M(x,y)$ to go from state (x,y) to state $(0,1)$ under the optimal policy satisfies

$$M(x,y) \leq \frac{1+p}{p^2}(c+y) + \frac{3}{2p^3}. \tag{18}$$

Proof. Note that

$$M(x,y) \leq \mathrm{E}[\hat{C}(x,y)], \tag{19}$$

where $\hat{C}(x,y)$ is the first passage cost under the policy in which the Reader samples in every slot. Starting from state (x,y) under the "always sample" policy, there is a geometric (p) number N of slots in which the system passes from states (x,y) up through $(x+N-1, y+N-1)$ until a memory update takes the system to state $(0, x+N)$. In the next slot, a cost $c+x+N$ is incurred and the system goes to either state $(0,1)$ with probability p or, with probability $1-p$, to $(1,1)$. In the latter case, the additional cost $\hat{C}(1,1)$ is incurred to reach $(0,1)$. We define the Bernoulli $(1-p)$ random variable $\overline{Z}$ such that $\overline{Z}=1$ if a memory update does *not* occur in state $(0, x+N)$. The cost expended to go from (x,y) to $(0,1)$ is then

$$\begin{aligned}\hat{C}(x,y) &= \sum_{j=y}^{y+N-1}(c+j) + (c+x+N) + \overline{Z}\hat{C}(1,1) \\ &= N(c+y) + (c+x) + \tfrac{N(N+1)}{2} + \overline{Z}\hat{C}(1,1). \end{aligned} \tag{20}$$

Taking expectation,

$$\mathrm{E}[\hat{C}(x,y)] = \frac{c+y}{p} + (c+x) + \frac{3-p}{2p^2} + \bar{p}\,\mathrm{E}[\hat{C}(1,1)]. \tag{21}$$

Evaluating (21) at $(x,y)=(1,1)$ yields

$$\mathrm{E}[\hat{C}(1,1)] = \frac{1}{p}\left[\left(\frac{1}{p}+1\right)(c+1) + \frac{3-p}{2p^2}\right]. \tag{22}$$

Combining (21) and (22) yields

$$\mathrm{E}[\hat{C}(x,y)] = \frac{c+y}{p} + (c+x) + \frac{1-p^2}{p^2}(c+1) + \frac{3-p}{2p^3}. \tag{23}$$

Since $x \leq y$ and $1 \leq y$ for any feasible state (x,y), we obtain

$$\begin{aligned}\mathrm{E}[\hat{C}(x,y)] &\leq \left(\frac{1}{p}+1\right)(c+y) + \frac{1-p^2}{p^2}(c+y) + \frac{3-p}{2p^3} \\ &\leq \frac{1+p}{p^2}(c+y) + \frac{3}{2p^3}. \end{aligned} \tag{24}$$

The claim then follows from (19). □

PROOF OF PROPOSITION 2

Proposition 2. If the α-optimal action is to sample in (x,y), then the α-optimal action is to sample in every (x,y') with $y' \geq y$.

Proof. For brevity, we'll use the following shorthand notation in the proof. For $w \leq v$, we define

$$\tilde{J}(u,v,w) = V(u,v) - V(u,w). \tag{25}$$

The monotonicity of the value function (Proposition 1) implies

$$\tilde{J}(u,v_1,w) \leq \tilde{J}(u,v_2,w) \quad \text{for all } u,\ w, \text{ and } v_1 \leq v_2. \tag{26}$$

For the rest of our discussion, we use the following form of discounted-cost Bellman's optimality equation with $c_\alpha = c/\alpha$:

$$\begin{aligned}V(x,y) = y + \alpha\min\{&pV(0,y+1) + \bar{p}V(x+1,y+1), \\ &c_\alpha + pV(0,x+1) + \bar{p}V(x+1,x+1)\}, \end{aligned} \tag{27}$$

Let $\hat{x} = x+1$ and $\hat{y} = y+1$. According to (27), the condition for the Reader to sample in (x,y) is

$$pV(0,\hat{y}) + \bar{p}V(\hat{x},\hat{y}) \geq c_\alpha + pV(0,\hat{x}) + \bar{p}V(\hat{x},\hat{x}). \tag{28}$$

Using the shorthand $\tilde{J}(u,v,w)$, the inequality (28) becomes

$$p\tilde{J}(0,\hat{y},\hat{x}) + \bar{p}\tilde{J}(\hat{x},\hat{y},\hat{x}) \geq c_\alpha. \tag{29}$$

Given that condition (29) holds, we examine the state $(x,\hat{y})$. The value function for this state is

$$\begin{aligned}V(x,\hat{y}) = \hat{y} + \alpha\min\{&pV(0,\hat{y}+1) + \bar{p}V(\hat{x},\hat{y}+1), \\ &c_\alpha + pV(0,\hat{x}) + \bar{p}V(\hat{x},\hat{x})\}. \end{aligned} \tag{30}$$

The condition for the Reader to sample in $(x,\hat{y})$ is

$$pV(0,\hat{y}+1) + \bar{p}V(\hat{x},\hat{y}+1) \geq c_\alpha + pV(0,\hat{x}) + \bar{p}V(\hat{x},\hat{x}), \tag{31}$$

or equivalently,

$$p\tilde{J}(0,\hat{y}+1,\hat{x}) + \bar{p}\tilde{J}(\hat{x},\hat{y}+1,\hat{x}) \geq c_\alpha. \tag{32}$$

Now we observe from the monotonicity property (26) and (29) that

$$\begin{aligned}p\tilde{J}(0,\hat{y}+1,\hat{x}) + \bar{p}\tilde{J}(\hat{x},\hat{y}+1,\hat{x}) &\geq p\tilde{J}(0,\hat{y},\hat{x}) + \bar{p}\tilde{J}(\hat{x},\hat{y},\hat{x}) \\ &\geq c_\alpha. \end{aligned} \tag{33}$$

Thus (32) holds, confirming that the Reader samples in state $(x,\hat{y})$. □

PROOF OF PROPOSITION 3

Proposition 3. (Concavity): $V(x,y)$ is concave in y. Specifically, for a fixed x, $V(x,y+1) - V(x,y)$ is non-increasing in y.

Proof. We want to show that for a fixed x, $V_n(x,i+1) - V_n(x,i) \geq V_n(x,i+2) - V_n(x,i+1)$, for every $i \in \mathbb{N}$. To achieve this, we focus on demonstrating the inequality:

$$V_n(x,i+2) + V_n(x,i) \leq 2V_n(x,i+1) \qquad \forall\ n, i. \tag{34}$$

The base case for $n=1$ is trivially satisfied, as $V_1(x,y) = y$. Now suppose (34) holds for $n = 1, 2\ldots k$ for every i. We will establish the validity of (34) under two scenarios, corresponding to the α-optimal action at stage $k+1$ being either to sample or idle in state $(x,i+1)$. First, let's consider the case where it is optimal to sample in $(x,i+1)$ at stage $k+1$. This implies that the value iteration function in this state satisfies:

$$\begin{aligned}V_{k+1}(x,i+1) = i+1+c+\alpha pV_k(0,x+1) \\ + \alpha\bar{p}V_k(x+1,x+1). \end{aligned} \tag{35}$$

Furthermore, leveraging Proposition 2, we deduce that sampling in $(x, i+1)$ is also the optimal action for state $(x, i+2)$ at stage $k+1$, resulting in:

$$V_{k+1}(x, i+2) = i + 2 + c + \alpha p V_k(0, x+1) \\ + \alpha \bar{p} V_k(x+1, x+1). \quad (36)$$

Notice that the value iteration function for (x, i) satisifies

$$V_{k+1}(x, i) \leq i + c + \alpha \left(p V_k(0, x+1) + \bar{p} V_k(x+1, x+1) \right). \quad (37)$$

Combining (35), (36), and (37), we establish:

$$V_{k+1}(x, i+2) + V_{k+1}(x, i) \leq 2V_{k+1}(x, i+1). \quad (38)$$

Let us now consider the situation where the α-optimal action is to stay idle in state $(x, i+1)$ at stage $k+1$. This implies that

$$V_{k+1}(x, i+1) = i + 1 + \alpha \left(p V_k(0, i+2) + \bar{p} V_k(x+1, i+2) \right). \quad (39)$$

Leveraging Proposition 2, we conclude that staying idle in $(x, i+1)$ is also the optimal action for state (x, i) at stage $k+1$, leading to:

$$V_{k+1}(x, i) = i + \alpha \left(p V_k(0, i+1) + \bar{p} V_k(x+1, i+1) \right). \quad (40)$$

The value iteration function for $(x, i+2)$ satisfies

$$V_{k+1}(x, i+2) \leq i + 2 + \alpha \left(p V_k(0, i+3) + \bar{p} V_k(x+1, i+3) \right). \quad (41)$$

Combining (40) and (41), we can demonstrate:

$$\begin{aligned} & V_{k+1}(x, i+2) + V_{k+1}(x, i) \\ & \leq 2(i+1) + \alpha [p \left(V_k(0, i+3) + V_k(0, i+1) \right) \\ & \qquad + \bar{p} \left(V_k(x+1, i+3) + V_k(x+1, i+1) \right)], \\ & \overset{(a)}{\leq} 2(i+1) + \alpha \left[2p V_k(0, i+2) + 2\bar{p} V_k(x+1, i+2) \right], \\ & = 2 \left(i + 1 + \alpha \left[p V_k(0, i+2) + \bar{p} V_k(x+1, i+2) \right) \right], \\ & \overset{(b)}{=} 2V_{k+1}(x, i+1), \end{aligned} \quad (42)$$

where (a) follows from induction hypothesis that $V_k(x, i+3) + V_k(x, i+1) \leq 2V_k(x, i+2)$, and (b) follows from (39). It follows from principle of mathematical induction that (34) holds for every n, and hence $V_n(x, y)$ is concave in y. As $\lim_{n \to \infty} V_n(x, y) = V(x, y)$, this implies that $V(x, y)$ is concave in y. □

PROOF OF PROPOSITION 4

Proposition 4. If the α-optimal action in state (x, y) is to idle, then the α-optimal action in states $(x+i, y+i), \forall i \geq 1$ is to stay idle.

Proof. For brevity, we'll use the following shorthand notation in the proof. For $w \leq v$, let

$$\tilde{J}_n(u, v, w) = V_n(u, v) - V_n(u, w). \quad (43)$$

We establish key properties of $\tilde{J}_n(u, v, w)$ to be utilized later in the proof.

1) Given $w \leq v$, Proposition 1 implies $V_n(u, v) \geq V_n(u, w) \geq 0$ and hence $\tilde{J}_n(u, v, w) \geq 0$.
2) If $v_2 \geq v_1$, and $w_2 \geq w_1$, it follows from concavity property (Proposition 3) that

$$V_n(u, v_2) - V_n(u, w_2) \leq V_n(u, v_1) - V_n(u, w_1) \quad (44)$$

and as a consequence,

$$\tilde{J}_n(u, v_2, w_2) \leq \tilde{J}_n(u, v_1, w_1), \forall u, \; w_1 \leq w_2, \text{ and } v_1 \leq v_2. \quad (45)$$

3) Let $\hat{u} = u + 1$, $\hat{v} = v + 1$ and $\hat{w} = w + 1$. Under the condition of not sampling in (u, v), it can be shown that

$$\tilde{J}_n(u, v, w) = v - w + \alpha (p \tilde{J}_{n-1}(0, \hat{v}, \hat{w}) + \bar{p} \tilde{J}_{n-1}(\hat{u}, \hat{v}, \hat{w})). \quad (46)$$

We now resume the proof of proposition. Letting $\hat{x} = x+1$ and $\hat{y} = y+1$ and $c_\alpha = c/\alpha$, we re-write the value iteration in state (x, y) given by (5) as:

$$V_{n+1}(x, y) = y + \alpha \min \{ p V_n(0, \hat{y}) + \bar{p} V_n(\hat{x}, \hat{y}), \\ c_\alpha + p V_n(0, \hat{x}) + \bar{p} V_n(\hat{x}, \hat{x}) \}, \quad (47)$$

Given that Reader doesn't sample in (x, y) implies that for all n, the terms inside the min function in (47) satisfy:

$$p V_n(0, \hat{y}) + \bar{p} V_n(\hat{x}, \hat{y}) \leq c_\alpha + p V_n(0, \hat{x}) + \bar{p} V_n(\hat{x}, \hat{x}). \quad (48)$$

Expressing inequality (48) in terms of $\tilde{J}_n(u, v, w)$, we get:

$$p \tilde{J}_n(0, \hat{y}, \hat{x}) + \bar{p} \tilde{J}_n(\hat{x}, \hat{y}, \hat{x}) \leq c_\alpha. \quad (49)$$

Given that (49) holds for every n, we examine state $(x+i, y+i)$. The value iteration expression at stage $n+1$ is given by:

$$\begin{aligned} & V_{n+1}(x+i, y+i) \\ & \quad = y + i + \alpha \min \{ p V_n(0, \hat{y}+i) + \bar{p} V_n(\hat{x}+i, \hat{y}+i), \\ & \qquad c_\alpha + p V_n(0, \hat{x}+i) + \bar{p} V_n(\hat{x}+i, \hat{x}+i) \}. \end{aligned} \quad (50)$$

To establish that the optimal action in state (x, y) being to stay idle implies the same for states $(x+i, y+i)$, we aim to show that the terms inside the min function in (50) satisfy:

$$p V_n(0, \hat{y}+i) + \bar{p} V_n(\hat{x}+i, \hat{y}+i) \leq c_\alpha + p V_n(0, \hat{x}+i) \\ + \bar{p} V_n(\hat{x}+i, \hat{x}+i). \quad (51)$$

or equivalently,

$$p \tilde{J}_n(0, \hat{y}+i, \hat{x}+i) + \bar{p} \tilde{J}_n(\hat{x}+i, \hat{y}+i, \hat{x}+i) \leq c_\alpha. \quad (52)$$

Given that (49) holds for all n, proving that (52) holds for all n is equivalent to showing that the LHS of (52) is less than LHS of (49). For that it is sufficient to show for all $n \geq 1$

$$I_1(n+1) = \tilde{J}_n(0, \hat{y}+i, \hat{x}+i) - \tilde{J}_n(0, \hat{y}, \hat{x}) \leq 0, \quad (53)$$

and

$$I_2(n+1) = \tilde{J}_n(\hat{x}+i, \hat{y}+i, \hat{x}+i) - \tilde{J}_n(\hat{x}, \hat{y}, \hat{x}) \leq 0. \quad (54)$$

With $i \geq 1$, it is clear that $\hat{y}+i \geq \hat{y}$ and $\hat{x}+i \geq \hat{x}$. Leveraging (45) , we conclude that $\tilde{J}_n(0, \hat{y}+i, \hat{x}+i) \leq \tilde{J}_n(0, \hat{y}, \hat{x})$, leading to $I_1 \leq 0$ for every n. We use inductive arguments to show that $I_2(n+1) \leq 0$. When $n=1$, we see that

$$\tilde{J}_1(u,v,w) = V_1(u,v) - V_1(u,w) = v - w. \quad (55)$$

This means that

$$I_2(2) = \tilde{J}_1(\hat{x}+i, \hat{y}+i, \hat{x}+i) - \tilde{J}_1(\hat{x}, \hat{y}, \hat{x}) = 0,$$

and hence the base case holds. Now assume that $I_2(n+1) \leq 0$ for $n = 1, \dots k-1$ for all $i \geq 0$. This implies:

$$I_2(k) = \tilde{J}_{k-1}(\hat{x}+i, \hat{y}+i, \hat{x}+i) - \tilde{J}_{k-1}(\hat{x}, \hat{y}, \hat{x}) \leq 0, \quad \forall i \geq 0. \quad (56)$$

We have established that $I_1(k) \leq 0$, implying that (53) holds at $n = k-1$. Combining this with (56), we conclude that both I_1 and I_2 hold at $n = k-1$. This, in turn, implies that (52) holds at $n = k-1$. Consequently, (51) holds at $n = k-1$. Therefore, the assumption $I_2(k) \leq 0$ for all $i \geq 0$ implies that the action that minimizes (50) at stage k is to stay idle in state $(x+i, y+i)$ for all $i \geq 1$.

Now, we need to demonstrate that:

$$I_2(k+1) = \tilde{J}_k(\hat{x}+i, \hat{y}+i, \hat{x}+i) - \tilde{J}_k(\hat{x}, \hat{y}, \hat{x}) \leq 0, \quad \forall i \geq 0. \quad (57)$$

Given that the assumption is to not sample in $(\hat{x}+i, \hat{y}+i)$ for $i \geq 0$ at stage k, this means that it is optimal to not sample in $(\hat{x}, \hat{y})$ (Proposition 2). Hence, employing Property (3) of $\tilde{J}_n(u,v,w)$, from (46), we have with $\hat{i} = i+1$,

$$\begin{aligned}
&\tilde{J}_k(\hat{x}+i, \hat{y}+i, \hat{x}+i) \\
&\quad = \hat{y} - \hat{x} + \alpha(p\tilde{J}_{k-1}(0, \hat{y}+\hat{i}, \hat{x}+\hat{i}) \\
&\qquad + \bar{p}\tilde{J}_{k-1}(\hat{x}+\hat{i}, \hat{y}+\hat{i}, \hat{x}+\hat{i})). \quad (58)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\tilde{J}_k(\hat{x}, \hat{y}, \hat{x}) &= \hat{y} - \hat{x} + \alpha(p\tilde{J}_{k-1}(0, \hat{y}+1, \hat{x}+1) \\
&\quad + \bar{p}\tilde{J}_{k-1}(\hat{x}+1, \hat{y}+1, \hat{x}+1)). \quad (59)
\end{aligned}$$

From (45), we can state that:

$$\tilde{J}_{k-1}(0, \hat{y}+\hat{i}, \hat{x}+\hat{i}) \leq \tilde{J}_{k-1}(0, \hat{y}+1, \hat{x}+1). \quad (60)$$

Additionally, it follows from (56),

$$\tilde{J}_{k-1}(\hat{x}+\hat{i}, \hat{y}+\hat{i}, \hat{x}+\hat{i}) \leq \tilde{J}_{k-1}(\hat{x}+1, \hat{y}+1, \hat{x}+1). \quad (61)$$

Based on (60) and (61), we observe that

$$\tilde{J}_k(\hat{x}+i, \hat{y}+i, \hat{x}+i) - \tilde{J}_k(\hat{x}, \hat{y}, \hat{x}) = I_2(k+1) \leq 0. \quad (62)$$

Thus, by induction, we establish that $I_2(n+1) \leq 0$ holds for all $n \geq 1$. □

PROOF OF PROPOSITION 6

Proposition 6. Defining $(0,1)$ as the reference state with $f(0,1) = 0$, the relative cost functions satisify:

(i)

$$f(0, Y_0+1) - f(0, Y_0) = 1. \quad (63)$$

Proof. Given that it is optimal to sample in $(0, Y_0)$, the relative-cost Bellman's equation in state $(0, Y_0)$ is given as

$$g + f(0, Y_0) = Y_0 + c + pf(0,1) + \bar{p}f(1,1). \quad (64)$$

The optimal action in $(0, Y_0+1)$ is also to sample (Proposition 2), and therefore, the relative-cost Bellman's equation in state $(0, Y_0+1)$ becomes

$$g + f(0, Y_0+1) = Y_0 + 1 + c + pf(0,1) + \bar{p}f(1,1). \quad (65)$$

It follows from (64) and (65) that $f(0, Y_0+1) - f(0, Y_0) = 1$. □

(ii) For any $x \geq 0$,

$$f(x, Y_0 - 1) = \frac{1}{p}(J_0 + \frac{\bar{p}}{p}) - 1, \quad (66)$$

where $J_0 = Y_0 - g + pf(0, Y_0)$.

Proof. For any $x \geq 0$, the optimal action in $(x, Y_0 - 1)$ is to idle, and the Bellman's equation (8) becomes

$$f(x, Y_0 - 1) = -g + Y_0 - 1 + pf(0, Y_0) + \bar{p}f(x+1, Y_0). \quad (67)$$

Let $J_0 = -g + Y_0 + pf(0, Y_0)$, we obtain

$$f(x, Y_0 - 1) = J_0 - 1 + \bar{p}f(x+1, Y_0). \quad (68)$$

Since the optimal action in $(x, Y_0 - 1)$ is to idle, then from Proposition 4, the optimal action in $x \geq 0$, $(x+1, Y_0)$, is to idle as well. The Bellman's equation (8) in $(x+1, Y_0)$ becomes

$$\begin{aligned}
&f(x+1, Y_0) \\
&= -g + Y_0 + pf(0, Y_0+1) + \bar{p}f(x+2, Y_0+1), \\
&\overset{(a)}{=} -g + Y_0 + p(1 + f(0, Y_0)) + \bar{p}f(x+2, Y_0+1), \\
&= J_0 + p + \bar{p}f(x+2, Y_0+1), \quad (69)
\end{aligned}$$

where (a) follows from Proposition 6(i). Substituting (69) into (68), we obtain

$$f(x, Y_0 - 1) = J_0(1+\bar{p}) - 1 + p\bar{p} + \bar{p}^2 f(x+2, Y_0+1). \quad (70)$$

Repeating this procedure n times yields

$$\begin{aligned}
f(x, Y_0 - 1) = J_0 \sum_{i=0}^{n} \bar{p}^i + p\bar{p} \sum_{i=1}^{n-1} (i+1)\bar{p}^i \\
+ \bar{p}^2 \sum_{i=0}^{n-2} (i+1)\bar{p}^i \\
+ \bar{p}^{n+1} f(x+n+1, Y_0+n) - 1, \quad (71)
\end{aligned}$$

and in the limit $n \to \infty$ we have

$$\begin{aligned}
f(0, Y_0 - 1) &= \frac{J_0}{1-\bar{p}} + \frac{p\bar{p}}{(1-\bar{p})^2} + \frac{\bar{p}^2}{(1-\bar{p})^2} - 1, \\
&= \frac{1}{p}(J_0 + \frac{\bar{p}}{p}) - 1. \quad (72)
\end{aligned}$$

Here, $\bar{p}^{n+1} f(x+n+1, Y_0+n) \to 0$ when $n \to \infty$ as $f(x+n+1, Y_0+n)$ is bounded. This bounding property is derived from Theorem 2, where it is established that $f(x+n+1, Y_0+n) \leq M(x+n+1, Y_0+n)$. Then it follows from (14),

$$f(x+n+1, Y_0+n) \leq \frac{1+p}{p^2}(c+Y_0+n) + \frac{3}{2p^3}. \quad (73)$$

$\square$

(iii) For every $y < Y_0$,

$$f(0, y) = f(1, y) = \cdots = f(y, y). \quad (74)$$

Proof. From the threshold structure of the optimal policy, the optimal action in $(x, Y_0 - 2)$ is to stay idle, and the relative-cost Bellman's equation (8) becomes

$$\begin{aligned} &f(0, Y_0 - 2) \\ &= -g + Y_0 - 2 + pf(0, Y_0 - 1) + \bar{p}f(1, Y_0 - 1), \\ &\overset{(a)}{=} -g + Y_0 - 2 + pf(0, Y_0 - 1) + \bar{p}f(0, Y_0 - 1), \\ &= -g + Y_0 - 2 + f(0, Y_0 - 1), \end{aligned} \quad (75)$$

where (a) follows from Proposition 6(ii) as $f(x, Y_0 - 1)$ is independent of x. This fact along with (75) implies that $f(x, Y_0 - 2)$ is also independent of x, and so $f(0, Y_0 - 2) = f(1, Y_0 - 2) \dots f(Y_0 - 2, Y_0 - 2)$. In fact this can be generalized such that $(x, Y_0 - k)$ with $x \geq 0$ and $k \in \{1, 2, \dots, Y_0 - 1\}$ is independent of x. $\square$

(iv) When $Y_0 > 1$,

$$f(0, Y_0) = Y_0 - g + c. \quad (76)$$

Proof. At $(0, Y_0)$ the Reader samples and the Bellman's equation (8) becomes

$$f(0, Y_0) = Y_0 - g + c + pf(0, 1) + \bar{p}f(1, 1). \quad (77)$$

When $Y_0 > 1$, we have from Proposition 6(iii), $f(0, 1) = f(1, 1)$, and since $f(0, 1) = 0$, it follows that

$$f(0, Y_0) = Y_0 - g + c. \quad (78)$$

$\square$

Proof of Lemma 1

Lemma 1. The average cost as a function of the threshold Y_0 is given by:

$$g_0(Y_0) = \frac{1}{2}\left(\frac{1}{p} + Y_0 + \frac{2cp + \bar{p}/p}{pY_0 + \bar{p}}\right). \quad (79)$$

Proof. We break down the proof into three parts. In the first and second parts, we derive analytical expressions for the optimal average cost when $Y_0 = 1$ and $Y_0 = 2$, respectively. In the third part, we focus on obtaining a general expression for the optimal average cost when $Y_0 > 2$. Surprisingly, we discover that the average cost equation as a function of Y_0 obtained in the third part is a general equation for any $Y_0 \geq 1$.

(*Part* 1): If $Y_0 = 1$, it is optimal to sample in $(0, 1)$. Thus Bellman's equation (8) yields

$$f(0, 1) = -g + 1 + c + pf(0, 1) + \bar{p}f(1, 1). \quad (80)$$

Defining $(0, 1)$ as the reference state with $f(0, 1) = 0$ yields $0 = -g + 1 + c + \bar{p}f(1, 1)$, or equivalently,

$$f(1, 1) = \frac{g - 1 - c}{\bar{p}}. \quad (81)$$

Now it is optimal to never sample in $f(1, 1)$. Then

$$\begin{aligned} f(1, 1) &= -g + 1 + pf(0, 2) + \bar{p}f(2, 2), \\ &\overset{(a)}{=} g + 1 + p(1 + f(0, 1)) + \bar{p}f(2, 2), \\ &= -g + 1 + p + \bar{p}f(2, 2), \end{aligned} \quad (82)$$

where (a) follows from Proposition 6(i). Since staying idle is the optimal action in $(1, 1)$, then Proposition 4 implies that staying idle is also the optimal action in state $(2, 2)$, and hence

$$\begin{aligned} f(2, 2) &= -g + 2 + pf(0, 3) + \bar{p}f(3, 3), \\ &= -g + 2 + p(2 + f(0, 1)) + \bar{p}f(3, 3), \\ &= -g + 2 + 2p + \bar{p}f(3, 3). \end{aligned} \quad (83)$$

Substituting $f(2, 2)$ obtained in (83) in (82), we obtain

$$f(1, 1) = -g(1 + \bar{p}) + 1 + p + 2\bar{p} + 2p\bar{p} + \bar{p}^2 f(3, 3). \quad (84)$$

Similar to above, we can obtain $f(3, 3)$ as

$$f(3, 3) = -g + 3 + 3p + \bar{p}f(4, 4). \quad (85)$$

Again substituting $f(3, 3)$ obtained in (85) in (84), we obtain

$$\begin{aligned} f(1, 1) = -g(1 + \bar{p} + \bar{p}^2) + 1 + p + 2\bar{p} + 2p\bar{p} + \\ 3\bar{p}^2 + 3p\bar{p}^2 + \bar{p}^3 f(4, 4). \end{aligned} \quad (86)$$

Repeating this n times we obtain

$$\begin{aligned} f(1, 1) = &-g(1 + \bar{p} + \bar{p}^2 \dots + \bar{p}^n) \\ &+ (1 + 2\bar{p} + 3\bar{p}^2 + \dots + (n + 1)\bar{p}^n) \\ &+ p(1 + 2\bar{p} + 3\bar{p}^2 + \dots + (n + 1)\bar{p}^n) \\ &+ \bar{p}^{n+1} f(n + 2, n + 2), \end{aligned} \quad (87)$$

and letting $n \to \infty$, we obtain

$$\begin{aligned} f(1, 1) &= \frac{-g}{1 - \bar{p}} + \frac{1}{(1 - \bar{p})^2} + \frac{p}{(1 - p)^2}, \\ &= \frac{-g}{p} + \frac{1}{p^2} + \frac{1}{p}, \\ &= \frac{-g + 1}{p} + \frac{1}{p^2}. \end{aligned} \quad (88)$$

Here, $\bar{p}^{n+1} f(n + 2, n + 2) \to 0$ when $n \to \infty$ as $f(n + 2, n + 2)$ is bounded. This bounding property is derived from Theorem 2, where it is established that $f(n + 2, n + 2) \leq M(n + 2, n + 2)$. Subsequently, (14) implies that

$$f(n + 2, n + 2) \leq \frac{1 + p}{p^2}(c + n + 2) + \frac{3}{2p^3}. \quad (89)$$

Now equating $f(1, 1)$ in (81) and (88), we obtain

$$g = \frac{1}{p} + cp. \quad (90)$$

(*Part* 2): If $Y_0 = 2$, then it is optimal to sample in $(0, 2)$. The realtive-cost Bellman's equation (8) is

$$\begin{aligned} f(0,2) &= -g + 2 + c + pf(0,1) + \bar{p}f(1,1), \\ &\overset{(a)}{=} -g + 2 + c, \end{aligned} \tag{91}$$

where (a) follows from Proposition 6(iii) which for $Y_0 = 2$ implies that $f(1,1) = f(0,1) = 0$. Now from (9), we have for $x = 1$, and $Y_0 = 2$,

$$\begin{aligned} f(0,1) &= \frac{1}{p}\left(J_0 + \frac{\bar{p}}{p}\right) - 1, \\ &= \frac{1}{p}\left(-g + 2 + pf(0,2) + \frac{\bar{p}}{p}\right) - 1. \end{aligned} \tag{92}$$

With $f(0,2)$ given by (91), we have

$$f(0,1) = \frac{1}{p}\left(-g + 2 + p(-g + 2 + c) + \frac{\bar{p}}{p}\right) - 1. \tag{93}$$

Equating $f(0,1) = 0$ gives

$$\begin{aligned} g &= \frac{1}{1+p}\left(2 + 2p + cp + \frac{\bar{p}}{p} - p\right), \\ &= \frac{1}{1+p}\left(1 + 1 + p + cp + \frac{\bar{p}}{p}\right), \\ &= \frac{1}{1+p}\left(1 + \frac{1}{p} + p(c+1)\right), \\ &= \frac{1}{p} + \frac{(c+1)p}{1+p}. \end{aligned} \tag{94}$$

(*Part* 3): Now consider the case when $Y_0 > 2$. Since it is optimal to not sample in state $(0, Y_0 - 2)$, the Bellman's equation for state $(0, Y_0 - 2)$ becomes

$$\begin{aligned} f(0, Y_0 - 2) &= Y_0 - 2 - g + pf(0, Y_0 - 1) + \bar{p}f(1, Y_0 - 1), \\ &\overset{(a)}{=} Y_0 - 2 - g + f(0, Y_0 - 1), \end{aligned} \tag{95}$$

where (a) follows from Proposition 6(iii). Moreover,

$$\begin{aligned} f(0, Y_0 - 3) &= Y_0 - 3 - g + pf(0, Y_0 - 2) + \bar{p}f(0, Y_0 - 2), \\ &= Y_0 - 3 - g + f(0, Y_0 - 2), \\ &= 2(Y_0 - g) - (2 + 3) + f(0, Y_0 - 1). \end{aligned} \tag{96}$$

Repeating this procedure k times yields,

$$\begin{aligned} &f(0, Y_0 - k) \\ &= (k-1)(Y_0 - g) - (2 + 3 + 4 + \cdots + k) + f(0, Y_0 - 1), \\ &= (k-1)(Y_0 - g) - \frac{k(k+1)}{2} + 1 + f(0, Y_0 - 1). \end{aligned} \tag{97}$$

Recalling $(0, 1)$ as the reference state with $f(0,1) = 0$, evaluating (97) at $k = Y_0 - 1$ yields

$$(Y_0 - 2)(Y_0 - g) = \frac{(Y_0 - 1)Y_0}{2} + 1 + f(0, Y_0 - 1), \tag{98}$$

From Proposition 6,

$$f(0, Y_0 - 1) = (1 + 1/p)(Y_0 - g) + c + \bar{p}/p^2 - 1. \tag{99}$$

Thus it follows from (98) that

$$(Y_0 - g)\left(\frac{1}{p} + Y_0 - 1\right) = \frac{(Y_0 - 1)Y_0}{2} - \frac{\bar{p}}{p^2} - c. \tag{100}$$

Rearranging to solve for g yields

$$\begin{aligned} g &= Y_0 - \frac{1}{Y_0 - 1 + 1/p}\left[\frac{Y_0(Y_0 - 1)}{2} - c - \frac{\bar{p}}{p^2}\right] \\ &= \frac{Y_0}{2} + \frac{\frac{Y_0(Y_0 - 1 + 1/p)}{2} - \left[\frac{Y_0(Y_0 - 1)}{2} - c - \frac{\bar{p}}{p^2}\right]}{Y_0 - 1 + 1/p} \\ &= \frac{Y_0}{2} + \frac{\frac{Y_0}{2p} + c + \frac{\bar{p}}{p^2}}{Y_0 - 1 + 1/p}. \end{aligned} \tag{101}$$

Recalling $-1 + 1/p = \bar{p}/p$, we obtain

$$g = \frac{Y_0}{2} + \frac{1}{2p} + \frac{1}{2}\frac{2cp + \bar{p}/p}{pY_0 + \bar{p}}. \tag{102}$$

Since (102) depends upon Y_0, we express it as

$$g_0(Y_0) = \frac{1}{2}\left(\frac{1}{p} + Y_0 + \frac{2cp + \bar{p}/p}{pY_0 + \bar{p}}\right). \tag{103}$$

Finally observe that even though (103) was derived for $Y_0 > 2$, we see that $g_0(1) = 1/p + cp$, where the RHS is same as RHS of (90), the average cost obtained separately at $Y_0 = 1$. Similarly, $g_0(2) = 1/p + (c+1)p/(1+p)$, where the RHS is same as RHS of (94), the average cost obtained separately at $Y_0 = 2$.

$\square$