

# Social Optimal Freshness in Multi-Source, Multi-Channel Systems via MDP

Shiksha Singhal, Veeraruna Kavitha and Vidya Shankar  
IEOR, Indian Institute of Technology Bombay, India

**Abstract**—Many systems necessitate frequent and consistent updates of a specific information. Often this information is updated regularly, where an old packet becomes completely obsolete in the presence of a new packet. In this context, we consider a system with multiple sources, each equipped with a storage buffer of size one, communicating to a common destination via  $d$  orthogonal channels. In each slot, the packets arrive at each source with certain probability and occupy the buffer (by discarding the old packet if any), and each transfer (to the destination) is successful with certain other probability. Thus in any slot, there are two (Age of Information) AoI-measures for each source: one corresponding to the information at the source itself and the other corresponding to the information of the same source available at the destination; some sources may not even have the packet to transmit.

The aim of the controller at the destination is to maintain the freshness of information of all the sources, to the best extent possible – it aims to design an optimal scheduling policy that assigns in each slot, a subset of sources with packets (at maximum  $d$ ) for transmission. This is achieved using an appropriate Markov Decision Process (MDP) framework, where the objective function is the sum of Average AoIs (AAoI) of all the sources. We derive a very simple stationary policy that is  $\epsilon$ -optimal – in any slot, order the sources with packets in the decreasing order of the differences in AoI at the destination and the source and choose the top sources for transmission. With moderate number of sources ( $< 30$ ), the AAoI reduces in the range of 30 – 90%.

## I. INTRODUCTION

With the advent of new technology and next generation networks that support smart applications, the need to continuously update information at centralised location from various sources becomes increasingly imperative; for example, Internet of Things (IoT), smart homes, environmental monitoring systems, on-the-road communication retrieval systems etc. The sources of information are required to transmit periodic status updates to their intended destinations (see [1]-[2]). A critical requirement for these services is ensuring that the information provided by the sources remains up-to-date at the destination, to the best extent possible; the quality of freshness of information is measured using “age of information” (AoI) which is the time elapsed since the time of generation of the latest available information [3]. However, this task is constrained by limited resources and the requirement coming from multiple sources. Thus we consider designing an optimal scheduling policy that optimises the sum of average age of information (AAoI) of all the sources.

Further, often in the systems that require regular updates of the same information, the old packet becomes obsolete once a new packet is available ([4]). Thus it is more appropriate to consider systems with at maximum one buffer storage, leading to lossy systems.

Motivated by the above factors, we consider a time-slotted system consisting of multiple sources (with single storage) that communicate to a common destination via multiple orthogonal channels. In each slot, the packets arrive at each source with certain probability, and each transfer (to the destination) is successful with certain other probability. Thus in any slot, there are two AoI-measures for each source: one corresponding to the information at the source itself and the other corresponding to the information of the same source available at the destination; some sources may not even have the packet to transmit (this happens when a source does not receive a new packet after its last successful transfer). At any time slot, these two AoI-measures corresponding to all the sources represent the state of the system. The goal is to design an optimal scheduling policy which determines the subset of sources for transfer in any time slot and which minimises the sum of the AAoI (at the destination) corresponding to all the sources.

**Related work:** The problem of minimising the age of information in such systems has been studied in [5]-[6] which focus on push-based communication where the sources decide when they want to send an update to the destination, and hence answer questions of optimal packet generation times. On the other hand, [7],[8],[9] and [10] focus on systems which implement pull-based communication, where the destination asks for data from the sources. Ours is a pull-based communication but with random packet arrivals to the sources and with some sources not having packets.

In [7], authors consider multiple independent sources providing status updates to a single destination via multiple orthogonal channels. The question here is similar: one needs to optimally choose a subset of sources to transmit, in each time slot. They assume the knowledge of (binary) channel conditions and choose for each channel one source among those that can communicate with the given channel in the given time slot. Further the sources always have information to transmit. In contrast, in our work the random channel conditions are unknown and we have more uncertainty in terms of packet availability at sources — it is not realistic to assume that the measurements are always available in all the time slots (measurement errors, transmission problems from the point of generation to the source itself, etc.), e.g., as in stock updates, sensor measurement, IoT, etc., which leads to a possibly stale (one or more slot old information) information even at the sources. Thus the policies in [7] consider only the AoIs at destination while we consider both sets of ages, including the ages at the sources.

The authors in [10] and [9] consider an infinite time

horizon problem and find optimal stationary randomised policies that are blind to the state of the system, i.e., without taking into account the ages of the packets at sources and the queue lengths (systems are not lossy) etc. The former work does not even consider the age of information at destination, while the latter work assumes perfect transmissions, i.e., the transmission probability equals one.

The study in [8] considers a framed-slot structure where the packets are generated in the beginning of each frame for each source and the scheduling policy allocates at maximum one source in each slot of the frame, based on the success of the previous transmission attempts of the same frame. Hence like in [7], the study in [8] also assumes the availability of packet at each source and in each frame and their policy is also similar; in each slot of the frame, select the source with highest age at destination and switching to a new source only when the packet is transmitted in the previous slot. This paper considers many more interesting aspects however, does not consider uncertainty related to packet availability at sources.

In contrast to the above strands of literature, we consider a system with unreliable and unknown channels (scheduling is blind to the channel conditions and hence scheduled transmission is successful with probability  $p$ ) and uncertainty in packet availability at sources. We however observe the AoIs at all the sources and destination for optimal source selection. Towards this, we derive an  $\epsilon$ -optimal scheduling policy and compare its performance with two policies inspired by the existing policies in literature; as already mentioned, none of the existing algorithms work under our assumptions (mainly uncertain packet availability at sources) and hence, we adapt them to our scenario and then compare with the proposed policy. In particular, we compare the proposed policy with: i) round-robin (RR) policy which chooses a subset of sources one after the other irrespective of the instantaneous ages and, ii) a partial information (PI) policy that chooses a subset of sources for transmission in any slot only based on instantaneous ages at the destination.

**Contributions:** The main contributions of this work are:

- We formulate the problem as an appropriate Markov Decision Process (MDP) and derive an  $\epsilon$ -optimal policy. Its performance approaches the performance of the optimal policy as  $p^2$  reduces to zero (Theorem 1). The policy is defined by a simple stationary rule, i.e., in any slot, order the sources with packets in the decreasing order of the differences in AoI at the destination and the source and choose the top sources for transmission,
- We demonstrate significant improvement (even up to 90%) with the moderate number of sources (Section V). More interestingly, the performance improvement is significant for all values of  $p$  (Figures 2 and 4) and,
- The performance of our policy starts matching with the existing policies when the number of sources is large (approximately  $> 100$ ).

## II. PROBLEM DEFINITION

Consider a system with a set  $\mathcal{N} = \{1, \dots, N\}$  of  $N$  sources, sending regular updates (packets) of a certain infor-

mation to a common destination via  $d$  orthogonal channels. We use a time-slotted system with  $T$  number of slots, where in each time slot of length  $\tau$ , each channel can be used by at most one source to transfer its packet to the destination. In every time slot, the packets arrive at any source  $n$  according to a Bernoulli process with probability  $q_n$  and a successful packet transfer to the destination happens with probability  $p$  for any source. All these events are independent of each other. The packets are identical but the transfer times may vary based on the random channel conditions – note here that the packet transfer times are geometric with parameter  $p$  for each channel. Each source has its own buffer with the storage capacity of one. *It is sufficient to consider storage capacity of one at every source, as the old packet becomes obsolete once a new packet arrives.* Our focus in this work is on measures related to the freshness of information available at the destination related to all the sources. Towards this, we measure the quality of information using a metric, called Age of Information (AoI) [3].

**Age of Information (AoI):** The age ( $H_n$ ) of information of source  $n$  at destination and at time-slot  $t$  is defined as the time elapsed since the last received packet at destination was generated, i.e.,

$$H_n(t) := t - r_n(t), \quad (1)$$

where  $r_n(t)$  is the time at which the last successfully received packet (at destination) before time  $t$ , is generated at source  $n$ .

**Average Age of Information (AAoI):** For any  $n \in \mathcal{N}$ , it is defined as below,

$$\bar{h}_n := \frac{\sum_{t=1}^T H_n(t)}{T}, \quad (2)$$

where  $T$  is the required time horizon. As already mentioned, we consider freshness of information in a lossy system, and our aim is to find an optimal source selection policy (for transmission in each slot) which minimises the sum of AAoI at the destination, from all the sources.

In this context, the state of the system at any time  $t$  can be represented by  $X(t) = (\mathbf{G}(t), \mathbf{H}(t))$  which is made up of two vectors with  $\mathbf{H}(t) = (H_1(t), \dots, H_N(t))$  as defined in (1) while  $\mathbf{G}(t) = (G_1(t), \dots, G_N(t))$  corresponds to the age of information at all the sources (see Figure 1). For example, if  $g_n(t) = k$  then the latest packet waiting at source  $n$  is generated in  $(t - k)^{th}$  slot.

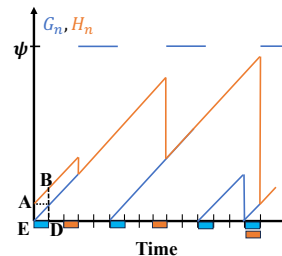


Fig. 1. Evolution of age at source and destination with time where blue and orange rectangles represent packet arrival to source  $n$  and successful packet transfer from source  $n$  to destination, respectively.

Another important point here is to observe that the old waiting packet is replaced with the new packet, if there is any arrival in any slot. Further, the age of any source  $n$  at time  $t$  is replaced with symbol  $\psi$ , if it has no packet; this can happen if its latest packet has been transmitted and there was no new packet arrival after that. *It is easy to see that  $G_n < H_n$  whenever source  $n$  has a packet.*

Note that the decision is simplified to selecting a subset of sources at the beginning of each time slot, instead of having to choose whether to continue with the current packet or choose a new packet (from the same/different source) as the transfer times are geometric and exhibit memory-less property. We model this as a finite horizon Markov Decision Process (MDP) whose ingredients we describe in the next section.

The aim of this paper is to derive optimal source scheduling policy which minimises the sum of the AAoI from all the sources at the destination (see (2)), i.e.,

$$\min \sum_{n \in \mathcal{N}} \bar{h}_n. \quad (3)$$

### III. MDP FORMULATION

In this section, we describe the MDP formulation and its ingredients.

**Decision Epochs:** The beginning of each time slot is considered as a decision epoch. We have  $1, \dots, T$  decision epochs with  $T < \infty$ .

**States:** As already defined, the state of the system at any time  $t$ ,  $X(t) = (\mathbf{G}(t), \mathbf{H}(t))$  which contains AoIs at all sources and destination at the beginning of time slot  $t$ . The state space is,

$$\mathcal{X} = \{x = (\mathbf{g}, \mathbf{h}) : g_i = \psi \text{ or } g_i < h_i\}.$$

**Actions:** The action at every decision epoch is to choose a subset of sources  $\mathbf{a}$ , whose packet is to be transferred. Let  $N_x := \sum_n \mathbf{1}_{\{g_n \neq \psi\}}$  be the number of sources that have packets to transmit. Then, the state dependent action space,

$$\mathcal{A}_x = 2^{S_x} \text{ with } S_x := \{n : g_n \neq \psi\}, N_x^d := \min\{N_x, d\}, \quad (4)$$

where  $2^{S_x}$  is a  $N_x^d$  sized subset of  $S_x$ , the set of sources with packets that can be transmitted. There are some observations related to this definition. Observe that there is only one action and  $\mathcal{A}_x = \{S_x\}$  for all the states with  $|S_x| = N_x^d \leq d$  and that no transmission is attempted if  $N_x^d = 0$ . Further, when  $N_x > d$ , any action  $\mathbf{a} \subset S_x$  with  $|\mathbf{a}| = d$ . In other words, we consider work conserving policies that facilitate transfer of all the available packets, however obviously constrained by capacity  $d$ .

**Cost:** Towards optimising (3), from (2) and (3), the appropriate instantaneous cost at time  $t$  when the state is  $x \in \mathcal{X}$  and action  $\mathbf{a} \in \mathcal{A}_x$  is chosen is given by,

$$c(x, \mathbf{a}) = \mathbb{E} \left[ \sum_{n=1}^N h_n | x, \mathbf{a} \right] \text{ for all } \mathbf{a} \in \mathcal{A}_x \text{ and all } x.$$

The above conditional expectation is the sum of  $N$  terms each of which is area under the age curve corresponding to one source (area under trapezium ABDE in Figure 1 is one such area corresponding to source  $n$ ) during a given time slot. Thus,

$$c(x, \mathbf{a}) = \sum_{n=1}^N (h_n \tau + 0.5 \tau^2) \text{ for all } \mathbf{a} \in \mathcal{A}_x \text{ and all } x.$$

Now optimising using the above instantaneous cost is equivalent to optimising using instantaneous costs, which equals  $c(x, \mathbf{a}) = \sum_{n=1}^N h_n \tau$  for all  $\mathbf{a} \in \mathcal{A}_x$  and all  $x$ , as  $\tau$  is an arbitrary constant. Thus, we set

$$c(x, \mathbf{a}) = \sum_{n=1}^N h_n \text{ for all } \mathbf{a} \in \mathcal{A}_x \text{ and all } x. \quad (5)$$

**Transition probabilities:** The system evolves to a new state  $X'$  when action  $\mathbf{a} \in \mathcal{A}_x$  is chosen in state  $x \in \mathcal{X}$  at any decision epoch. This transition depends on the transfer status (if any, i.e., if  $N_x^d > 0$ ) and packet arrival status, both in the previous slot. Let  $\mathcal{C} \subset \mathcal{N}$  be the set of sources which generated new packets in the duration between previous and current decision epoch. Let  $X(t-1) = x$  and  $\mathbf{A}(t-1) = \mathbf{a}$ .

Now, if the packet of sources  $W \subset \mathbf{a}$  is successfully transferred then the state evolves as below: for any  $n$ ,

$$H'_n = \begin{cases} h_n + 1 & \text{for all } n \notin W, \\ g_n + 1 & \text{else, i.e., if } n \in W. \end{cases} \quad (6)$$

$$G'_n = \begin{cases} g_n + 1 & \text{for all } n \notin \mathcal{C} \text{ and } n \notin W, \\ 0 & \text{for all } n \in \mathcal{C}, \\ \psi & \text{else, i.e., if } n \notin \mathcal{C} \text{ and } n \in W. \end{cases} \quad (7)$$

Observe that the age of packet at source  $a$ ,  $G'_a$  is set to  $\psi$  when it does not receive a new packet, as the latest packet with it is just transferred. The transition of the above type (specified by  $W$  and  $\mathcal{C}$ ) happens with probability,

$$\mathbb{P}(X(t) = X' | x, \mathbf{a}) = p^{|\mathcal{W}|} (1-p)^{|\mathcal{A}| - |\mathcal{W}|} \prod_{n \in \mathcal{N} \setminus \mathcal{C}} (1 - q_n) \prod_{n \in \mathcal{C}} q_n. \quad (8)$$

Thus with all the ingredients defined, the optimisation of the term in (3) is equivalent to solving the following MDP where,

$$V^*(x) = \min_{\pi = (d_1, \dots, d_{T-1})} \mathbb{E}_{\pi, x} \left[ \sum_{t=1}^T c(X(t), \mathbf{A}(t)) \right].$$

### IV. NEAR OPTIMAL POLICY

The aim of this section is to derive near optimal or  $\epsilon$ -optimal policy. It is well known in the MDP literature that the optimal policy can be non-stationary (i.e., the decision rules are different across time slots) for finite horizon problems [11]. However, interestingly our  $\epsilon$ -optimal policy turns out to be stationary. In particular, we consider the following special stationary policy constructed using the differences between AoIs at sources and destinations as defined below:

$$\pi^\Delta = (d_1^\Delta, \dots, d_{T-1}^\Delta) \text{ where } d_t^\Delta = d^\Delta \text{ for all } t, \text{ and}$$

$$d^\Delta(x) = \arg \min_{\mathbf{a} \in \mathcal{A}_x} \sum_{i=1}^{N_x} (g_{a_i} - h_{a_i}), \quad (9)$$

as already explained (in Section III) there is only one action when  $N_x \leq d$ . Next, we present our main result which upper bounds the performance of  $\pi^\Delta$  policy with respect to the optimal value (proof in Appendix).

*Theorem 1: There exists a sequence of non-negative functions  $\{\mathcal{Z}_t(x)\}$  for each  $t \leq T$  such that the stage-wise optimal value function(s)  $\{V_t^*(x)\}$  are related to the corresponding value function(s)  $\{V_t^\Delta(x)\}$  under policy  $\pi^\Delta$  of (9) as below:*

$$V_t^\Delta(x) - V_t^*(x) = p \cdot p_d \mathcal{Z}_t(x) \text{ for each } t, x, \text{ with,} \quad (10)$$

$$p_d := 1 - (1-p)^d = p C_p, \quad C_p = \left( \sum_{l=0}^{d-1} (-1)^l \binom{d}{l+1} p^l \right).$$

Further there exist two constants  $\{D_1(t)\}$  and  $\{D_2(t)\}$  independent of  $p$  such that,

$$|\mathcal{Z}_t(x)| \leq D_1(T-t) \|x\|_\infty + D_2(T-t) \text{ with} \quad (11)$$

$$\|x\|_\infty := \max_{n \in \mathcal{N}} \{h_n + 1\}.$$

**Remarks:** From the above theorem (see (10)-(11)) the difference,

$$V^\Delta(x) - V^*(x) \leq p \cdot p_d (D_2(1) + \|x\|_\infty D_1(1)) \text{ for each } x,$$

where constants  $D_1(1), D_2(1)$  are independent of  $p$ . Thus clearly as  $p \rightarrow 0$ , we have  $p \cdot p_d = p^2 C_p \rightarrow 0$  and thus the objective function  $V^\Delta(x)$  evaluated under stationary policy  $\pi^\Delta$  of (9) approaches the value function  $V^*(x)$ . In fact this approach is uniform in all  $x \in \mathcal{X}_B$  for any  $B < \infty$ , where  $\mathcal{X}_B := \{x : g_n = \psi \text{ or } g_n < h_n \text{ and } h_n \leq B \forall n\}$ .

Thus  $\pi^\Delta$  is  $\epsilon$ -optimal when initial condition belongs to  $\mathcal{X}_B$  and when  $p\sqrt{C_p} < \sqrt{\epsilon}/(D_1(1)(B+1) + D_2(1))$ . This establishes near-optimality of  $\pi^\Delta$  when  $p^2$  is sufficiently small (observe  $C_p \leq d$  for all  $p$ ).

Observe that the above approximation is good as  $p^2 \rightarrow 0$  and this already suggests that even for moderate  $p$  (and for which  $p^2$  is negligible) one can have a good approximation. In fact, more interestingly we observe via simulations in the next section that the  $\pi^\Delta$  policy performs significantly well even for values of  $p$  close to one, in comparison with the policies known in literature.

## V. SIMULATIONS

In this section, we compare the performance of our proposed policy with the policies in the existing literature. To the best of our knowledge, it appears that there is no work in the literature that considers our scenario, i.e., unreliability at packet generation and transfer; further the decision maker does not have access to the channel conditions.

The authors in [7] consider generic binary channels (for example, a kind of Markovian channel) but assume the

availability of the packets at all the sources, in each time slot; they also assume the knowledge of the channel conditions before taking a decision. Their policy is to select a subset of sources (among the sources with good channel conditions) with the highest AoIs at destination for transfer, in each time slot.

On the other hand, [8] considers a framed slotted structure, with each source having a packet at the beginning of the frame. They propose a modified Robin Round (greedy) policy, which order the sources in the decreasing order of AoIs at the destination in the first frame. Henceafter, they follow the Round Robin policy, i.e., chooses sources one after the other, with switching only once the packet is transferred. The specific choice in the first frame and the Round Robin henceafter ensures that the source selected in any time slot is the one with the highest AoI at the destination.

We compare our  $\epsilon$ -optimal policy (9) with the one obtained by adapting the policies in [7] and [8] to our scenario, i.e., to the case with unreliable packet generation — basically, at any time instance we choose a subset of sources based on the order of the AoIs at the destination among the sources with packets to transfer. Since this policy does not consider the AoI at sources, we refer it as partial information (PI) policy.

We also compare our policy (referred to as  $\epsilon$ -O policy) with a complete blind policy, which chooses sources for transfer one after the other, irrespective of the AoIs and the transfer status of the packets. We refer it as Round Robin (RR) policy.

In Figures 2-3, we consider an example with packet arrival probabilities  $q_n = .5$  for all sources  $n \in \mathcal{N}$  and  $d = 1$ . We consider comparison across different number of sources or different values of transmission probability  $p$ , in this study. One can make several observations as below,

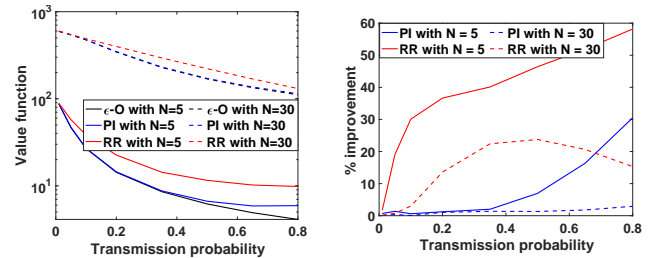


Fig. 2. **Packet arrivals**,  $q_n = 0.5 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus transmission probability  $p$  with  $d = 1$

(a) In Figure 2, we consider comparison across various  $p$  for two different values of  $N$ , the solid lines are for  $N = 5$ , while the dash lines are for  $N = 30$ . Different policies are represented by different colours. It is easy to see from the figures that the improvement in value function as compared to the PI policy is almost negligible for smaller values of  $p$  while it is large and up to 55% for higher values of  $p$ . Further, even though for  $\epsilon$ -O policy, theoretical guarantees have been established for smaller values of  $p$ , it still performs better than RR and PI for higher values of  $p$ .

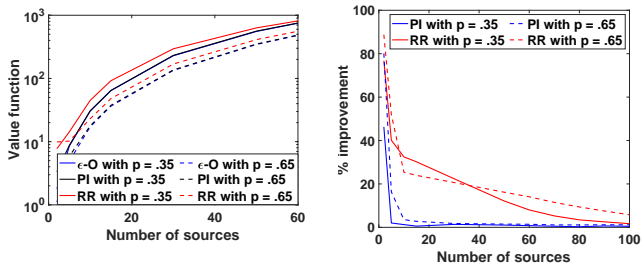


Fig. 3. **Packet arrivals**,  $q_n = 0.5 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus number of sources  $N$  with  $d = 1$

(b) In Figure 3, we consider comparison across various number of sources for two different values of  $p$ , the solid lines are for  $p = 0.65$ , while the dash lines are for  $p = 0.35$ . Different policies are again represented by different colours.

Clearly,  $\epsilon$ -O policy and PI policy outperforms the RR policy (given by red curves); this is true even in Figure 2. However, more interestingly, as the number of sources increases, the performance of RR policy also approaches the performance of the remaining two policies (one can see the convergence towards the right of the Figure 3, and the percentage improvement is less than 5%, in both the cases, when the number of sources is near 100). In fact, we observe this in many other examples. *Thus one can use this much simplified blind RR policy when the number of sources is extremely large.*

(c) In all the case studies, the  $\epsilon$ -O policy outperforms the other two. The percentage improvement over the partial information or PI policy is up to 80% (higher when the number of sources is small) and that against blind RR policy is even higher.

Next, we consider an example with even rarer packet arrivals, where  $q_n = .1$  for all  $n \in \mathcal{N}$  in Figures 4-5.

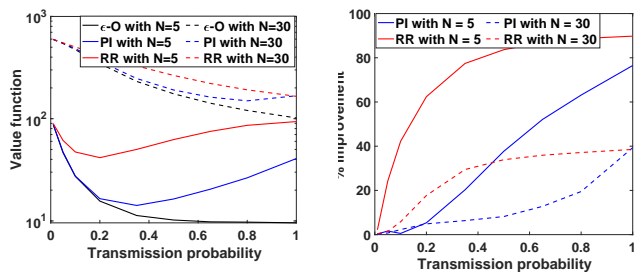


Fig. 4. **Rare packet arrivals**,  $q_n = 0.1 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus transmission probability  $p$  with  $d = 1$

In contrast to the previous example, we see even bigger improvements for this rare arrival case study; one can anticipate this as the AoIs at sources also convey significant information in such cases and we managed to derive a simple stationary policy that uses both sets of AoI; the increase in the complexity is minimal (one needs to order according

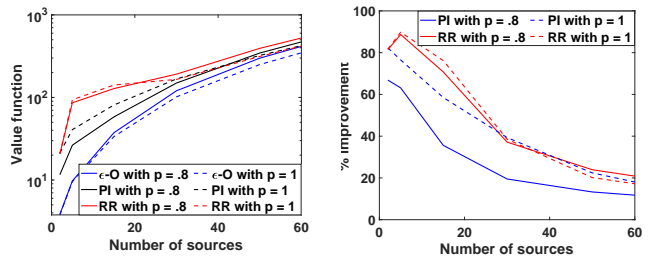


Fig. 5. **Rare packet arrivals**,  $q_n = 0.1 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus number of sources  $N$  with  $d = 1$

the differences in AoIs at the source and the destination as compared to ordering according to AoIs just at the destination), yet  $\epsilon$ -O policy effectively provides significant improvement.

Similar trends are observed for other case studies with multiple channels (with  $d = 3$ ) in Figures 6-9; now the improvements are even higher, for example in the right sub-figure of Figure 6 when  $p$  is close to 1 and the number of sources  $N = 30$ , the improvement is up to 90%, even for the case with reasonable packet arrival rates, i.e., with  $q_n = .5$ .

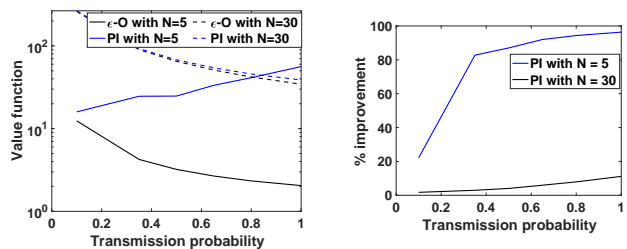


Fig. 6. **Packet arrivals**,  $q_n = 0.5 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus transmission probability  $p$  with  $d = 3$

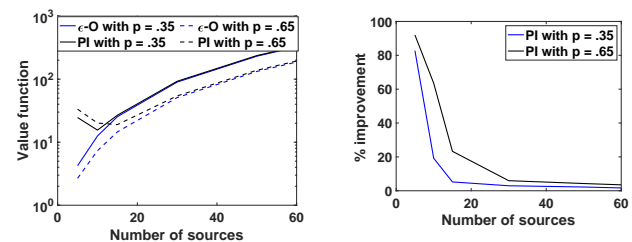


Fig. 7. **Packet arrivals**,  $q_n = 0.5 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus number of sources  $N$  with  $d = 3$

## VI. CONCLUSION

We consider a system with multiple sources trying to transmit information packets to a common destination via multiple orthogonal channels. Each source has its own buffer with single storage capacity. The packets arrive to each of the sources according to a Bernoulli process while the transfer times are geometric. We formulate it as a finite horizon Markov Decision Process (MDP) and derive a near optimal policy. Interestingly, the derived policy (defined only in terms



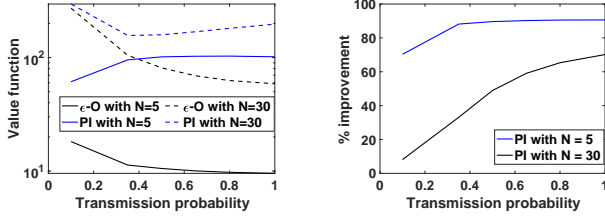


Fig. 8. **Rare packet arrivals**,  $q_n = 0.1 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus transmission probability  $p$  with  $d = 3$

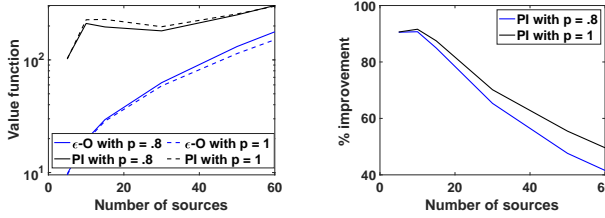


Fig. 9. **Rare packet arrivals**,  $q_n = 0.1 \forall n$ : Value function (on log scale) in left sub-figure and percentage improvement in the right figure versus number of sources  $N$  with  $d = 3$

of the differences in AoIs at the sources and the destination) is stationary and further does not depend on the packet arrival rates to any of the sources; this additional computational advantage is mainly due to the fact that the policy is near optimal. We demonstrate the superiority of the proposed policy by comparing it with the adaptations of the existing policies in literature for our case with unreliable packet generation, through numerical experiments. There are several future directions here, for example, to study the source selection policy for multiple sources transmitting packets to multiple destinations via multiple channels.

## REFERENCES

- [1] Ahmed M Bedewy, Yin Sun, and Ness B Shroff. Optimizing data freshness, throughput, and delay in multi-server information-update systems. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 2569–2573. IEEE, 2016.
- [2] Peter Corke, Tim Wark, Raja Jurdak, Wen Hu, Philip Valencia, and Darren Moore. Environmental wireless sensor networks. *Proceedings of the IEEE*, 98(11):1903–1917, 2010.
- [3] Sanjit Kaul, Marco Gruteser, Vinuth Rai, and John Kenney. Minimizing age of information in vehicular networks. In *2011 8th Annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks*, pages 350–358. IEEE, 2011.
- [4] Veeraruna Kavitha and Eitan Altman. Controlling packet drops to improve freshness of information. In *Network Games, Control and Optimization: 10th International Conference, NetGCoop 2020, France, September 22–24, 2021, Proceedings 10*, pages 60–77. Springer, 2021.
- [5] Yin Sun, Elif Uysal-Biyikoglu, Roy D Yates, C Emre Koksall, and Ness B Shroff. Update or wait: How to keep your data fresh. *IEEE Transactions on Information Theory*, 63(11):7492–7508, 2017.
- [6] Roy D Yates and Sanjit Kaul. Real-time status updating: Multiple sources. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 2666–2670. IEEE, 2012.
- [7] Vishrant Tripathi and Sharayu Moharir. Age of information in multi-source systems. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.
- [8] Igor Kadota, Abhishek Sinha, Elif Uysal-Biyikoglu, Rahul Singh, and Eytan Modiano. Scheduling policies for minimizing age of

information in broadcast wireless networks. *IEEE/ACM Transactions on Networking*, 26(6):2637–2650, 2018.

- [9] Yu-Pin Hsu, Eytan Modiano, and Lingjie Duan. Age of information: Design and analysis of optimal scheduling algorithms. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 561–565. IEEE, 2017.
- [10] Igor Kadota and Eytan Modiano. Minimizing the age of information in wireless networks with stochastic arrivals. In *Proceedings of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 221–230, 2019.
- [11] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

## APPENDIX

**Proof of Theorem 1:** We directly compute the difference between the objective function under the optimal policy (i.e, value function) and that under the policy (9). We derive it by solving the following DP equations using backward recursion.

$$\begin{aligned} V_T^*(x) &= c(x) \text{ for all } x \text{ and for any } t < T, \forall x, \\ V_t^*(x) &= \min_{\mathbf{a} \in \mathcal{A}_x} [c(x, \mathbf{a}) + \mathbb{E}[V_{t+1}^*(X')|x, \mathbf{a}]]. \end{aligned} \quad (12)$$

Recall now  $\mathbf{a} = (a_1, a_2, \dots, a_{N^d})$  is the subset of sources that attempted transition at state  $x$  in the previous slot.

**At terminal epoch, i.e., at  $t = T$ :** There are no further transmission attempts and hence from (5) and (12),

$$V_T^*(x) = c(x) = \sum_{n \in \mathcal{N}} h_n \text{ for all } x. \quad (13)$$

**At decision epoch  $t = T - 1$ :** Towards computing value function, one needs to derive Q-functions corresponding to each possible action, and then

$$V_{T-1}^*(x) = \min_{\mathbf{a} \in \mathcal{A}_x} Q_{T-1}(x, \mathbf{a}). \quad (14)$$

Consider any  $x$  and any  $\mathbf{a} \in \mathcal{A}_x$ . We begin with computing  $Q(x, \mathbf{a})$ . Towards that, let  $H'_n$  be the age of the packet corresponding to source  $n$  at destination in time slot  $T$  when sources in subset  $\mathbf{a}$  are chosen for transfer – observe that  $H'_n = h_n + 1$  for all  $n \notin \mathbf{a}$  and otherwise  $H'_n = g_n + 1$  or  $h_n + 1$  depending upon the success of the transfer of packet from source  $n$ . Thus from (12)-(14),

$$\begin{aligned} Q_{T-1}(x, \mathbf{a}) &= \sum_{n \in \mathcal{N}} h_n + \mathbb{E} \left[ \sum_{n \in \mathcal{N}} H'_n | x, \mathbf{a} \right], \\ &= 2 \sum_{n \in \mathcal{N}} h_n + N + pl(x, \mathbf{a}) \text{ with,} \\ l(x, \mathbf{a}) &:= \sum_{i=1}^{N^d} (g_{a_i} - h_{a_i}), \end{aligned}$$

where the last equality follows from Lemma 1. Hence, using (9) with  $l(x) = l^\Delta(x) := \min_{\mathbf{a} \in \mathcal{A}_x} l(x, \mathbf{a})$ , we have for all  $x$

$$V_{T-1}^*(x) = 2 \sum_{n \in \mathcal{N}} h_n + N + pl^\Delta(x) = V_{T-1}^\Delta(x). \quad (15)$$

Observe that the last equality follows as the decisions while defining  $V_{T-1}^*$  are the same as in that in  $\pi^\Delta$  of (9).

**At decision epoch  $t = T - 2$ :** We again compute the Q-functions using (15) and with  $\{H'_n\}, \{G'_n\}, X'$  now representing the quantities at  $T - 1$ . For any  $x$  and  $\mathbf{a} \in \mathcal{A}_x$ ,

$$\begin{aligned}
Q_{T-2}(x, \mathbf{a}) &= \sum_{n \in \mathcal{N}} h_n + \mathbb{E} \left[ \left( 2 \sum_{n \in \mathcal{N}} H'_n + N + pl^\Delta(X') \right) \middle| x, \mathbf{a} \right], \\
&= 3 \sum_{n \in \mathcal{N}} h_n + 3N + 2pl(x, \mathbf{a}) + p\mathbb{E}[l^\Delta(X') | x, \mathbf{a}], \\
&= 3 \sum_{n \in \mathcal{N}} h_n + 3N + 2pl(x, \mathbf{a}) + p(1 - p_x^d)\mathcal{U}(x) \\
&\quad + p \cdot p_d \mathcal{V}(x, \mathbf{a}),
\end{aligned}$$

where the last equality follows from simple algebra and Lemma 2 with  $\mathcal{U}(x)$  and  $\mathcal{V}(x, \mathbf{a})$  as defined in Lemma 2. Thus for any  $x$ ,

$$\begin{aligned}
V_{T-2}^*(x) &= 3 \sum_{n \in \mathcal{N}} h_n + 3N + p(1 - p_x^d)\mathcal{U}(x) \\
&\quad + p \min_{\mathbf{a} \in \mathcal{A}_x} \left[ 2l(x, \mathbf{a}) + p_d \mathcal{V}(x, \mathbf{a}) \right].
\end{aligned}$$

Similarly, under  $\pi^\Delta$  of (9), with  $\mathbf{a}^\Delta = d^\Delta(x)$ ,  $V_{T-2}^\Delta(x) = Q_{T-2}(x, \mathbf{a}^\Delta)$ . Now with  $l^*(x) := l(x, \mathbf{a}^*)$  and  $\mathbf{a}^* := \arg \min_{\mathbf{a} \in \mathcal{A}_x} [2l(x, \mathbf{a}) + p_d \mathcal{V}(x, \mathbf{a})]$ ,

$$\begin{aligned}
\left| V_{T-2}^\Delta(x) - V_{T-2}^*(x) \right| &= V_{T-2}^\Delta(x) - V_{T-2}^*(x), \\
&= 2p[l^\Delta(x) - l^*(x)] \\
&\quad + p \cdot p_d (\mathcal{V}(x, \mathbf{a}^\Delta) - \mathcal{V}(x, \mathbf{a}^*)), \\
&\leq p \cdot p_d (\mathcal{V}(x, \mathbf{a}^\Delta) - \mathcal{V}(x, \mathbf{a}^*)), \\
&\leq p \cdot p_d (2d\|x\|_\infty),
\end{aligned}$$

where the last inequality again follows from Lemma 2. Hence at  $t = T - 2$ ,  $V_t^*(x) = V_t^\Delta(x) - p \cdot p_d \mathcal{Z}_t(x)$  where the difference function is bounded as  $0 \leq \mathcal{Z}_t(x) \leq 2d\|x\|_\infty$ . Next, assume that the value function and the objective under policy  $\pi^\Delta$  has the following form for any  $k \geq 2$ ,

$$V_{T-k}^*(x) = V_{T-k}^\Delta(x) - p \cdot p_d \mathcal{Z}_{T-k}(x), \quad (16)$$

$$V_{T-k}^\Delta(x) = (k+1) \sum_{n \in \mathcal{N}} h_n + K(k) + p\tilde{V}_{T-k}(x) \text{ with } (17)$$

$$K(k) = kN + K(k-1), \quad (18)$$

$$\tilde{V}_{T-k}(x) = \mathcal{F}_{T-k}(x) + kl^\Delta(x) \quad (19)$$

for some appropriate functions,  $\mathcal{F}_{T-k}$  and  $\mathcal{Z}_{T-k}$  which can be upper bounded as below,

$$\mathcal{F}_{T-k}(x) \leq C_1(k)\|x\|_\infty + C_2(k), \quad (20)$$

$$\mathcal{Z}_{T-k}(x) \leq D_1(k)\|x\|_\infty + D_2(k), \text{ with, } \quad (21)$$

$$C_1(k) = (1 + p_d)C_1(k-1) + 2(k-1)d, \quad (22)$$

$$C_2(k) = (1 + p_d)[C_1(k-1) + C_2(k-1)], \quad (23)$$

$$D_1(k) = (1 + p_d)D_1(k-1) + 2C_1(k-1) + 2(k-1)d, \quad (24)$$

$$D_2(k) = (1 + p_d)[D_1(k-1) + D_2(k-1)] + 2[C_1(k-1) + C_2(k-1)]. \quad (25)$$

Observe that for  $k = 2$ , (16)-(25) are satisfied with,

$$\mathcal{F}_{T-2}(x) = (1 - p_x^d)\mathcal{U}(x) + p_d \mathcal{V}(x, \mathbf{a}^\Delta), \quad K(1) = N,$$

$$\mathcal{Z}_{T-2}(x) = \frac{2}{p_d} [l^\Delta(x) - l^*(x)] + \mathcal{V}(x, \mathbf{a}^\Delta) - \mathcal{V}(x, \mathbf{a}^*).$$

and  $\mathcal{F}_{T-2}(x) \leq (1 + p_d)d\|x\|_\infty$  (with  $C_1(2) = (1 + p_d)d$  and  $C_2(2) = 0$ ) by Lemma 2, which satisfies (20); it is

already proved that the bound on  $\mathcal{Z}_{T-2}(x)$  (with  $D_1(2) = 2d$  and  $D_2(2) = 0$ ) satisfies (21). By backward mathematical induction, it suffices to show that the value function and the objective have the above form at  $t = T - (k + 1)$ . For any  $x$  and  $\mathbf{a} \in \mathcal{A}_x$  using Lemma 1,

$$\begin{aligned}
Q_{T-(k+1)}(x, \mathbf{a}) &= \sum_{n \in \mathcal{N}} h_n + \mathbb{E} \left[ (k+1) \sum_{n \in \mathcal{N}} H'_n + K(k) + p\tilde{V}_{T-k}(X') \right. \\
&\quad \left. - p \cdot p_d \mathcal{Z}_{T-k}(X') \middle| x, \mathbf{a} \right], \\
&= (k+2) \sum_{n \in \mathcal{N}} h_n + K(k+1) + (k+1)pl(x, \mathbf{a}) \\
&\quad + p\mathbb{E} \left[ \tilde{V}_{T-k}(X') - p_d \mathcal{Z}_{T-k}(X') \middle| x, \mathbf{a} \right], \quad (26)
\end{aligned}$$

with  $K(k+1) = (k+1)N + K(k)$  which satisfies same equation as (18), but at  $k+1$ . Further using (19) and Lemma 2 (with  $\mathcal{U}(x)$  and  $\mathcal{V}(x, \mathbf{a})$  defined there and bounded by  $d\|x\|_\infty$ ) the last but one term of (26),

$$\begin{aligned}
\mathbb{E} \left[ \tilde{V}_{T-k}(X') | x, \mathbf{a} \right] &= \mathbb{E}[\mathcal{F}_{T-k}(X') | x, \mathbf{a}] + k[(1 - p_x^d)\mathcal{U}(x) + p_d \mathcal{V}(x, \mathbf{a})].
\end{aligned}$$

Substituting the above in (26), we obtain,

$$\begin{aligned}
Q_{T-(k+1)}(x, \mathbf{a}) &= (k+2) \sum_{n \in \mathcal{N}} h_n + K(k+1) + (k+1)pl(x, \mathbf{a}) \\
&\quad + p\mathbb{E}[\mathcal{F}_{T-k}(X') | x, \mathbf{a}] + kp[(1 - p_x^d)\mathcal{U}(x) + p_d \mathcal{V}(x, \mathbf{a})] \\
&\quad - p \cdot p_d \mathbb{E}[\mathcal{Z}_{T-k}(X') | x, \mathbf{a}]. \quad (27)
\end{aligned}$$

Conditioning on  $\mathbb{S}$  (flag indicating at least one successful packet transfer) and following steps exactly as in Lemma 2, we get functions (which depend on time)  $\mathcal{G}_{T-k}(x)$ ,  $\mathcal{G}'_{T-k}(x)$ ,  $\mathcal{H}_{T-k}(x, \mathbf{a})$ , and  $\mathcal{H}'_{T-k}(x, \mathbf{a})$ , one can show that,

$$\mathbb{E}[\mathcal{F}_{T-k}(X') | x, \mathbf{a}] = (1 - p_x^d)\mathcal{G}_{T-k}(x) + p_d \mathcal{H}_{T-k}(x, \mathbf{a}),$$

$$\mathbb{E}[\mathcal{Z}_{T-k}(X') | x, \mathbf{a}] = (1 - p_x^d)\mathcal{G}'_{T-k}(x) + p_d \mathcal{H}'_{T-k}(x, \mathbf{a}).$$

Further observe  $\|X'\|_\infty \leq \|x\|_\infty + 1$  a.s. irrespective of  $\mathbf{a}$  and hence,  $\mathcal{F}_{T-k}(X') \leq C_1(k)(\|x\|_\infty + 1) + C_2(k)$  using (20). Conditioning on  $\mathbb{S} = 0$ , we have,

$$\mathbb{E}[\mathcal{F}_{T-k}(X') | x, \mathbf{a}, \mathbb{S} = 0] = \mathcal{G}_{T-k}(x).$$

Thus we have  $|\mathcal{G}_{T-k}(x)| \leq C_1(k)(\|x\|_\infty + 1) + C_2(k)$ , from above and using (20). Similar argument follows and the upper bound for  $\mathcal{H}_{T-k}(x, \mathbf{a})$  matches with that on  $\mathcal{G}_{T-k}(x)$ ; further  $\mathcal{G}'_{T-k}(x)$  and  $\mathcal{H}'_{T-k}(x, \mathbf{a})$  can be upper bounded with  $D_1(k)(\|x\|_\infty + 1) + D_2(k)$ . Hence from (27),

$$\begin{aligned}
V_{T-(k+1)}^*(x) &= (k+2) \sum_{n \in \mathcal{N}} h_n + K(k+1) + p(1 - p_x^d)\mathcal{G}_{T-k}(x) \\
&\quad + kp(1 - p_x^d)\mathcal{U}(x) - p \cdot p_d(1 - p_x^d)\mathcal{G}'_{T-k}(x) \\
&\quad + p \min_{\mathbf{a} \in \mathcal{A}_x} \left[ (k+1)l(x, \mathbf{a}) + p_d \mathcal{H}_{T-k}(x, \mathbf{a}) \right. \\
&\quad \left. - (p_d)^2 \mathcal{H}'_{T-k}(x, \mathbf{a}) + kp_d \mathcal{V}(x, \mathbf{a}) \right]. \quad (28)
\end{aligned}$$

In similar lines using (17)-(19),

$$\begin{aligned}
V_{T-(k+1)}^\Delta(x) &= (k+2) \sum_{n \in \mathcal{N}} h_n + K(k+1) + p(1-p_x^d) \mathcal{G}_{T-k}(x) \\
&\quad + kp(1-p_x^d) \mathcal{U}(x) + (k+1)pl^\Delta(x) \\
&\quad + p \cdot p_d \left[ \mathcal{H}_{T-k}(x, \mathbf{a}^\Delta) + k\mathcal{V}(x, \mathbf{a}^\Delta) \right]. \tag{29}
\end{aligned}$$

Now, the difference between the objective under policy of (9) and value function is bounded as below,

$$\begin{aligned}
p \cdot p_d \mathcal{Z}_{T-(k+1)}(x) &:= V_{T-(k+1)}^\Delta(x) - V_{T-(k+1)}^*(x), \\
&= p \cdot p_d(1-p_x^d) \mathcal{G}'_{T-k}(x) + (k+1)p[l^\Delta(x) - l^*(x)] \\
&\quad + p \cdot p_d \left[ \mathcal{H}_{T-k}(x, \mathbf{a}^\Delta) - \mathcal{H}_{T-k}(x, \mathbf{a}^*) \right] \\
&\quad + k[\mathcal{V}(x, \mathbf{a}^\Delta) - \mathcal{V}(x, \mathbf{a}^*)] + p_d \mathcal{H}'_{T-k}(x, \mathbf{a}^*), \\
&\leq p \cdot p_d(1-p_x^d) \mathcal{G}'_{T-k}(x) \\
&\quad + p \cdot p_d \left[ \mathcal{H}_{T-k}(x, \mathbf{a}^\Delta) - \mathcal{H}_{T-k}(x, \mathbf{a}^*) \right] \\
&\quad + k[\mathcal{V}(x, \mathbf{a}^\Delta) - \mathcal{V}(x, \mathbf{a}^*)] + p_d \mathcal{H}'_{T-k}(x, \mathbf{a}^*), \\
&\leq p \cdot p_d \left[ (1+p_d)[D_1(k)(\|x\|_\infty + 1) + D_2(k)] \right. \\
&\quad \left. + 2[C_1(k)(\|x\|_\infty + 1) + C_2(k)] + 2kd\|x\|_\infty \right], \\
&= p \cdot p_d \left[ [(1+p_d)D_1(k) + 2C_1(k) + 2kd]\|x\|_\infty \right. \\
&\quad \left. + (1+p_d)[D_1(k) + D_2(k)] + 2[C_1(k) + C_2(k)] \right], \tag{30}
\end{aligned}$$

which thus satisfies (16) and (21) using the recursive constants defined as in (22)-(25). Comparing (29) with (17)-(19), the function  $\mathcal{F}_{T-(k+1)}$  can be identified and bounded as below,

$$\begin{aligned}
\mathcal{F}_{T-(k+1)}(x) &= (1-p_x^d) \mathcal{G}_{T-k}(x) + k(1-p_x^d) \mathcal{U}(x) \\
&\quad + p_d \left[ \mathcal{H}_{T-k}(x, \mathbf{a}^\Delta) + k\mathcal{V}(x, \mathbf{a}^\Delta) \right], \\
&\leq [C_1(k)(\|x\|_\infty + 1) + C_2(k)] + k(1+p_x^d)d\|x\|_\infty \\
&\quad + p_d [C_1(k)(\|x\|_\infty + 1) + C_2(k)], \\
&= [(1+p_d)C_1(k) + kd(1+p_x^d)]\|x\|_\infty + (1+p_d)[C_1(k) + C_2(k)], \\
&\leq [(1+p_d)C_1(k) + 2kd]\|x\|_\infty + (1+p_d)[C_1(k) + C_2(k)], \\
&= C_1(k+1)\|x\|_\infty + C_2(k+1).
\end{aligned}$$

In the above we used (22) and (23) and then (20) is satisfied. Further, since it is a finite horizon problem, the above constants are bounded and since  $p \leq 1$ , the bounds can be independent of  $p$ . ■

*Lemma 1: Given any  $x$  and  $\mathbf{a} \in \mathcal{A}_x$ , we have*

$$\mathbb{E} \left[ \sum_{n \in \mathcal{N}} H'_n | x, \mathbf{a} \right] = \sum_{n \in \mathcal{N}} h_n + N + pl(x, \mathbf{a}).$$

**Proof:** The above conditional expectation equals,

$$\begin{aligned}
&= \sum_{n \in \mathbf{a}} [p(g_n + 1) + (1-p)(h_n + 1)] + \sum_{n \notin \mathbf{a}} (h_n + 1), \\
&= \sum_{n \in \mathcal{N}} h_n + N + pl(x, \mathbf{a}). \quad \blacksquare
\end{aligned}$$

*Lemma 2: Consider any  $x$  and  $\mathbf{a} \in \mathcal{A}_x$ . For any time  $t$  conditioned on  $X_t = x$  and  $\mathcal{A}_t = \mathbf{a}$ , and with  $X'$  representing the quantities at  $t+1$ , we have the following: there exists two non-negative-valued bounded functions  $\mathcal{U}$  and  $\mathcal{V}$  such that the former depends only on  $x$  (independent of  $t, \mathbf{a}$ ) while the latter depends on both  $x$  and  $\mathbf{a}$  (independent of  $t$ ) and one can express the conditional expectation as,*

$$\mathbb{E} \left[ l(X') | x, \mathbf{a} \right] = (1-p_x^d) \mathcal{U}(x) + p_d \mathcal{V}(x, \mathbf{a}).$$

where  $p_x^d = 1 - (1-p)^{N_x^d}$ . and  $p_d = 1 - (1-p)^d$ . Further irrespective of  $\mathbf{a}$

$$|\mathcal{U}(x)| \leq d\|x\|_\infty \text{ and } |\mathcal{V}(x, \mathbf{a})| \leq d\|x\|_\infty.$$

**Proof:** (i) Let  $\mathbb{S}$  be a flag indicating at least one successful packet transfer. Then,

$$\begin{aligned}
\mathbb{E} \left[ l(X') | x, \mathbf{a} \right] &= (1-p_x^d) \mathbb{E} \left[ l(X') | x, \mathbf{a}, \mathbb{S} = 0 \right] \\
&\quad + p_x^d \mathbb{E} \left[ l(X') | x, \mathbf{a}, \mathbb{S} = 1 \right].
\end{aligned}$$

Now, it is easy to observe that the state transitions corresponding to first term are independent of the action chosen as  $G'_{\hat{a}} = 1_{\{\mathbb{R}_{\hat{a}}=0\}}(g_{\hat{a}} + 1)$  and  $H'_{\hat{a}} = h_{\hat{a}} + 1$  for all  $\hat{a} \in S_X'$ , where  $\mathbb{R}_{\hat{a}}$  be the indicator of new packet arrivals at source  $\hat{a}$  (observe here that with  $\mathbb{S} = 0$ ,  $S_X \subset S_X'$  and  $\mathbb{R}_{\hat{a}} = 1$  for  $\hat{a} \in S_X' \setminus S_X$  for almost all  $X'$ ). Thus for some appropriate function  $\mathcal{U}$  of  $x$  (alone), the first term equals,

$$(1-p_x^d) \mathbb{E} \left[ l(X') | x, \mathbf{a}, \mathbb{S} = 0 \right] = (1-p_x^d) \mathcal{U}(x).$$

When conditioned on  $\mathbb{S} = 1$ , it is clear that transitions depend not only on  $x$  but also on  $\mathbf{a}$ . Hence, there exists a function  $\mathcal{V}(\cdot)$  of  $x$  and  $\mathbf{a}$  such that,

$$p_x^d \mathbb{E} \left[ l(X') | x, \mathbf{a}, \mathbb{S} = 1 \right] = p_d \frac{p_x^d}{p_d} \mathbb{E} \left[ l(X') | x, \mathbf{a}, \mathbb{S} = 1 \right] = p_d \mathcal{V}(x, \mathbf{a}).$$

Now we are left to derive the upper bound on functions  $\mathcal{U}$  and  $\mathcal{V}$ . Consider any source  $\hat{a}$  with packet in the new state  $X'$ . For this source, when conditioned on  $\mathbb{S} = 0$ , the  $G_{\hat{a}} - H_{\hat{a}}$  term has the following probabilistic description (see (6)-(8)),

$$[(G'_{\hat{a}} - H'_{\hat{a}}) | x, \mathbf{a}, \mathbb{S} = 0] = \begin{cases} g_{\hat{a}} - h_{\hat{a}} & \text{w.p. } (1 - q_{\hat{a}}), \\ (-1 - h_{\hat{a}}) & \text{w.p. } q_{\hat{a}}. \end{cases}$$

In either case, i.e., almost surely,  $(G'_{\hat{a}} - H'_{\hat{a}})$  is upper bounded by  $h_{\hat{a}} + 1$  (recall  $g_{\hat{a}} < h_{\hat{a}}$ ) and hence the absolute value  $|l(X')|$  is a.s. upper bounded by  $d\|x\|_\infty$ . Therefore  $|\mathcal{U}|$  is also upper bounded by the same quantity. When conditioned on  $\mathbb{S} = 1$ , the  $G_{\hat{a}} - H_{\hat{a}}$  terms have the following probabilistic description, and result follows by similar logic,

$$\begin{aligned}
&[(G'_{\hat{a}} - H'_{\hat{a}}) | x, \mathbf{a}, \mathbb{S} = 1] \\
&= \begin{cases} g_{\hat{a}} - h_{\hat{a}} & \text{if } \mathbb{R}_{\hat{a}} = 0 \text{ and } \hat{a} \notin \mathbf{a}, \text{ w.p. } (1 - q_{\hat{a}}), \\ (-1 - h_{\hat{a}}) & \text{if } \mathbb{R}_{\hat{a}} = 1 \text{ and } \hat{a} \notin \mathbf{a}, \text{ w.p. } q_{\hat{a}}, \\ (-g_{\hat{a}} - 1) & \text{if } \mathbb{R}_{\hat{a}} = 1 \text{ and } \hat{a} \in \mathbf{a}, \text{ w.p. } q_{\hat{a}}. \end{cases} \quad \blacksquare
\end{aligned}$$