

# Sequential Problems in Decentralized Detection with Communication

Ashutosh Nayyar, *Student Member, IEEE*, and Demosthenis Teneketzis, *Fellow, IEEE*

**Abstract**—Sequential problems in decentralized detection are considered. One or more peripheral sensors can make repeated noisy observations of a binary hypothesis on the state of the environment. At each time, a peripheral sensor has to decide whether to stop and send a final binary message to a coordinating sensor/fusion center or to continue taking measurements. Every time a peripheral sensor postpones its final message to the coordinating sensor, it incurs a penalty. The coordinating sensor's operation is explored under two different scenarios. In the first scenario, the coordinating sensor waits to receive the final message from each of the peripheral sensors and then starts taking measurements of its own. Every time the coordinating sensor takes a measurement, it incurs a penalty/measurement cost. The coordinating sensor is then faced with a stopping problem on whether to stop and declare a decision on the hypothesis or to continue taking measurements. In the second scenario, the coordinating sensor starts taking measurements from the beginning, without waiting for the final messages from the peripheral sensors. It is then faced with a different stopping problem. At any time, the coordinating sensor has to decide whether to stop and declare a decision on the hypothesis or to continue to take more measurements and wait for the peripheral sensors to send their final message. Parametric characterization of optimal policies for the peripheral and the coordinating sensors are obtained under both scenarios. A sequential methodology for finding the optimal policies is presented.

## I. INTRODUCTION

Decentralized detection problems are motivated by applications in large scale decentralized systems such as sensor networks and surveillance networks. In such networks, sensors receive different information about the environment but share a common objective, for example to detect the presence of a target in a surveillance area or to detect the occurrence of a fault in a power system. Sensors may be allowed to communicate but they are constrained to exchange only a limited amount of information because of energy constraints, communication constraints, data storage and data processing constraints etc.

Decentralized detection problems may be static or sequential. In static problems, sensors make a fixed number of observations about a hypothesis on the state of the environment which is modeled as a random variable  $H$ . Sensors may transmit a single message (a quantized version of their observations) to a fusion center which makes a final decision on  $H$ . Such problems have been extensively studied since their initial

formulation in [1] (See the surveys in [2], [3] and references therein). In most such formulations, it has been shown that person-by-person optimal decision rules (as defined in [4]) for a binary hypothesis detection problem are characterized by thresholds on the likelihood ratio (or equivalently on the posterior belief on the hypothesis). Under certain conditions such as large number of identical sensors, it has been shown that it is optimal to use identical quantization rules at all sensors ([5], [6]). A related information-theoretic formulation with constraints on communication from a sensor to a fusion center appears in [7].

In sequential problems, the number of observations taken by the sensors is not fixed a priori. Two distinct scenarios have been considered for sequential problems. In one scenario, at each time instant local/peripheral sensors send a message about their observations to a fusion center/coordinator. At each time instant, the fusion center decides whether to receive more messages (and incur more measurement costs) or to stop and declare a decision on the hypothesis. Thus the fusion center is faced with an optimal stopping problem whereas the peripheral sensors are faced with a quantization problem. The case where peripheral sensors can only use their current observation and possibly all past transmissions of all sensors to decide what message to send to the fusion center has been studied in [8]. Asymptotic results where the cost of taking a measurement approaches zero or the probability of making a wrong decision approaches zero have been found in the case when sensors remember their past observations as well ([9], [10]).

A second scenario for sequential detection problems may be motivated by situations where continuous communication with a fusion center is too costly. For example, remote wireless sensors with low energy budget may not be capable of continuous communication with a fusion center. In this scenario, each sensor locally decides when to stop taking more measurements and only sends a final message to a fusion center. Each sensor incurs an operational cost proportional to the number of measurements it takes and/or a penalty for delaying its final decision. Thus, in this scenario, each local/peripheral sensor is faced with an optimal stopping problem. Several possible cases can be considered in this scenario depending on what the fusion center is allowed to do:

*Case I:* The fusion center waits to receive the final messages from all sensors and then combines them according to a fixed rule (say a majority rule) to produce a final decision on the hypothesis. A version of this problem (called the decentralized Wald problem) was formulated in [11] and it was shown that at each time instant, optimal policies for the peripheral sensors are described by two thresholds on

A. Nayyar is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48109 USA e-mail: (anayyar@umich.edu).

D. Teneketzis is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, 48109 USA e-mail: (teneket@eecs.umich.edu).

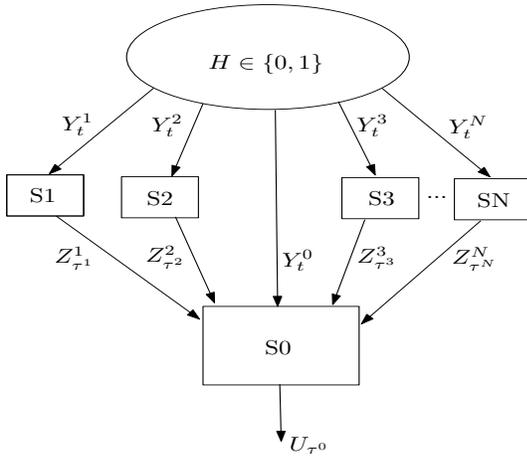


Fig. 1. Decentralized Detection with  $N$  Peripheral Sensors sending their final messages  $Z_{\tau^i}^i$  at stopping time  $\tau^i$  to a Coordinating Sensor ( $S_0$ ) that makes the final decision at stopping time  $\tau^0$

their posterior beliefs. The computation of these thresholds requires solution of two coupled sets of dynamic programming equations. Similar results were obtained in a continuous time setting in [12].

*Case II:* After receiving the final messages from all the sensors, the fusion center has the option to make observations of its own. It is possible that the fusion center has some uncertainty about the hypothesis after receiving all sensor messages possibly because of conflicting messages from different sensors. In such a situation, the fusion center may choose to make its own observations. The fusion center then has a stopping problem - it has to decide when to stop taking observations and what final decision to make. Like the sensors, the fusion center also incurs a penalty for delaying its final decision. A cost depending on the final decision on the hypothesis and the true value of the hypothesis is incurred in the end.

*Case III:* In a sensor network, the role of fusion center may be performed by one of the sensors itself. In this case, the fusion center (that is, the coordinating sensor) is making observations of its own from the beginning *and* receiving final messages from other sensors. Like the peripheral sensors, the coordinating sensor has a stopping problem. At any time, the coordinating sensor uses its own observations and the messages it has received so far to make a decision on whether to stop and declare a final decision on the hypothesis or continue to take more observations and wait for messages from other sensors that have not yet sent a final message. As in Case II, each sensor (peripheral sensors and the coordinating sensor) incurs a penalty for delaying its final message/decision, and a cost depending on the coordinating sensor's final decision on the hypothesis and the true value of the hypothesis is incurred in the end.

The problems formulated in this paper are motivated by the scenarios in Cases II and III above. We first consider a simple two sensor (a peripheral and a coordinating sensor) version of the problems in Case II and Case III and obtain a parametric characterization of optimal policies. We prove that at each

time instant, an optimal policy of the peripheral sensor is characterized by at most 4 thresholds on its posterior belief on the hypothesis; an optimal policy of the coordinating sensor is characterized by 2 thresholds (on its own posterior belief) that depend on the messages received from the peripheral sensor. This characterization differs from the classical two threshold characterization found in the centralized and the decentralized Wald problems ([13], [11]). A fundamental feature of our problems in contrast to the problems in Case I is that peripheral sensors convey information to the coordinating sensor not only by the choice of the final message but also through the time at which this message is sent. Thus, at times when the peripheral sensor chooses not to send a decision, it still conveys information to the fusion center who observes that the final message was not sent at this time. This feature is the main conceptual reason for the difference between our parametric characterizations and the classical two-threshold characterization.

Our parametric characterization significantly reduces the complexity of finding optimal strategies. Instead of optimizing over the space of policies (which are functions from observations to actions), the optimization can be restricted to space of thresholds. The computation of globally optimal thresholds is still a difficult problem. We present a sequential methodology that decomposes the overall optimization problem into several smaller problems that may be solved to determine the optimal thresholds at each time instant. We then extend our results to the general problems with multiple peripheral sensors described in Cases II and III. We show that qualitative properties of the optimal policies of the peripheral sensors and the coordinating sensor are the same as in the two-sensor problem.

We also extend our parametric characterizations to infinite horizon problems. We show by an example that the expected stationarity of optimal policies in infinite horizon problems with stationary observation statistics does not hold in our decentralized framework.

The rest of the paper is organized as follows. In Section II, we formulate two versions of our problem with two observers/sensors (corresponding to Cases II and III). We obtain qualitative results on the nature of optimal policies for the two sensors in Sections III and IV. We present a sequential methodology for computing optimal policies in Section V. In Section VI, we extend our qualitative results to infinite horizon analogues of our problem. A generalization to more than binary communication alphabet is presented in Section VII. We extend our results to multiple sensor (more than 2) problems in section VIII. Finally, we conclude in Section IX.

*Notation:* Throughout this paper,  $X_{1:t}$  refers to the sequence  $X_1, X_2, \dots, X_t$ . Subscripts are used as time index and the superscripts are used as the index of the sensor. We use capital letters to denote random variable and the corresponding lower case letters for their realizations.  $P(\cdot)$  denotes probability of an event and  $\mathbb{E}(\cdot)$  denotes expectation of a random variable; for a collection of decision functions  $\Gamma$ ,  $P^\Gamma(\cdot)$  and  $\mathbb{E}^\Gamma(\cdot)$  are used to indicate that the probability/expectation are defined for the fixed choice of  $\Gamma$ .

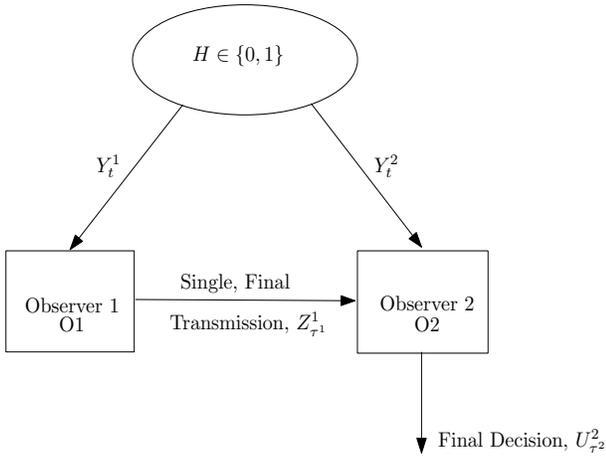


Fig. 2. Decentralized Detection with two observers

## II. PROBLEM FORMULATION

### A. The Model

Consider a binary hypothesis problem where the true hypothesis is modeled as a random variable  $H$  taking values 0 or 1 with known prior probabilities:

$$P(H = 0) = p_0; \quad P(H = 1) = 1 - p_0$$

Consider two observers: Observer 1 (O1) and Observer 2 (O2). We assume that each observer can make noisy observations of the true hypothesis. Conditioned on the hypothesis  $H$ , the following statements are assumed to be true:

1. The observation of the  $i^{\text{th}}$  observer at time  $t$ , ( $Y_t^i$ ) (taking values in the set  $\mathcal{Y}^i$ ), either has a discrete distribution ( $P_t^i(\cdot|H)$ ) or admits a probability density function ( $f_t^i(\cdot|H)$ ).
2. Observations of the  $i^{\text{th}}$  observer at different time instants are conditionally independent given  $H$ .
3. The observation sequences at the two observers are conditionally independent given  $H$ .

Observer 1 observes the measurement process  $Y_t^1, t = 1, 2, \dots$ . At any time  $t$ , after having observed the sequence of observations  $Y_{1:t}^1$ , observer 1 can decide either to stop and send a binary message 0 or 1 to observer 2 or to postpone its decision and get another measurement. Each time observer 1 postpones its decision, a cost of  $c^1$  is incurred. (The cost  $c^1$  incorporates the additional cost of taking a new measurement, the energy cost of staying on for another time step and/or a penalty for delaying the decision.) Note that observer 1 transmits only a single final binary message (0 or 1) to observer 2.

The decision of observer 1 at time  $t$  is based on the entire sequence of observations till that time, in other words, observer 1 has perfect recall. Thus, we have that

$$Z_t^1 = \gamma_t^1(Y_{1:t}^1), \quad (1)$$

where  $Z_t^1$  is observer 1's message at time  $t$  to observer 2 and  $\gamma_t^1$  is the *decision-function* used by O1 at time  $t$ .  $Z_t^1$  belongs to the set  $\{0, 1, b\}$ , where we use  $b$  for blank message, that is, no transmission. The sequence of functions  $\gamma_t^1, t = 1, 2, \dots$ , constitutes the *policy* of observer 1. Let  $\tau^1$  be the stopping

time when observer sends a final message to observer 2, that is,

$$\tau^1 = \min\{t : Z_t^1 \in \{0, 1\}\} \quad (2)$$

We allow two possibilities for the operation of observer 2. *Case A:* In this case, observer 2 does not take any measurements until it receives the final message from O1. After receiving observer 1's final message, observer 2 can decide either to stop and declare a decision on the hypothesis or to start taking measurements on its own. After observer 2 has made  $k$  measurements ( $k = 1, 2, \dots$ ), it can decide to stop and declare a final decision on the hypothesis or take a new measurement. Each time observer 2 decides to take another measurement it incurs a cost  $c^2$ . Whenever observer 2 makes a final decision  $U \in \{0, 1\}$  on the hypothesis, it incurs a cost  $J(U, H)$ . As in the case of observer 1, we assume observer 2 has perfect recall. Let  $U_k^2 \in \{0, 1, N\}$  be the decision made by observer 2 after receiving  $\tau^1$  messages ( $Z_{1:\tau^1}^1$ ) from observer 1 and subsequently making  $k$  observations ( $Y_{1:k}^2$ ) of its own, (where we use  $N$  for a null decision, that is, a decision to continue taking measurements). Thus,

$$U_k^2 = \gamma_k^2(Y_{1:k}^2, Z_{1:\tau^1}^1), \quad (3)$$

for  $k = 0, 1, 2, \dots$  (Note that after  $\tau^1$ , the time index  $k$  starts from 0 for O2 where  $k = 0$  means that O2 makes a final decision without making any measurements of its own). The message sequence  $Z_{1:\tau^1}^1$  is a sequence of  $\tau^1 - 1$  blank messages followed by  $Z_{\tau^1}^1 = 0$  or 1. The sequence of decision-functions  $\gamma_k^2, k = 0, 1, 2, \dots$  constitutes the *policy* of observer 2. We define  $\tau^2$  to be the number of measurements taken by observer 2 before announcing its final decision on the hypothesis, that is,

$$\tau^2 = \min\{k : U_k^2 \in \{0, 1\}\} \quad (4)$$

*Case B:* In this case, O2 starts taking measurements at time  $t = 1$  without waiting for O1 to send a final message. At time  $t = 1, 2, \dots$ , we have the following time-ordering of the two observers' observations and decisions:

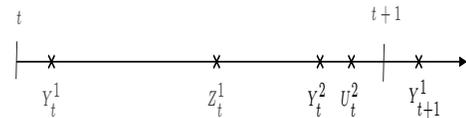


Fig. 3. Time ordering in Case B (Problem P2)

Thus, observer 2's decision at time  $t$  can be described as:

$$U_t^2 = \gamma_t^2(Y_{1:t}^2, Z_{1:t}^1) \quad (5)$$

where  $U_t^2 \in \{0, 1, N\}$ . This decision is a function of the observations made at O2 ( $Y_{1:t}^2$ ) and the messages received from O1 ( $Z_{1:t}^1$ ). The message sequence  $Z_{1:t}^1$  could be a sequence of  $t$  blank messages received from O1 or  $k$  blanks ( $k < t$ ) followed by a 0 or 1. Let  $\tau^2$  be the stopping time when observer 2 announces its final decision on the hypothesis, that is,

$$\tau^2 = \min\{t : U_t^2 \in \{0, 1\}\} \quad (6)$$

Note that we allow O2 to declare a final decision without getting the final message from O1. Also, O1 does not know

whether O2 has stopped or not, that is, there is no feedback from O2 to O1. As in Case A, a penalty of  $c^2$  is incurred every time O2 decides to postpone its final decision and a terminal cost of  $J(U, H)$  is incurred when O2 makes its final decision  $U \in \{0, 1\}$ .

In both the cases above, we assume that the cost parameters  $c^1, c^2$  are finite positive numbers and  $J(U, H)$  is non-negative and bounded by a finite constant  $L$  for all  $U$  and  $H$ . Moreover, we assume that cost of an error in the final decision is more than cost of a correct decision, that is,  $J(0, 1) > J(1, 1)$  and  $J(1, 0) > J(0, 0)$ . We can now formulate an optimization problem for each of the two cases above.

1) *Problem P1*: We consider a finite horizon  $T^1$  for observer 1. That is, if the observer 1 has not sent its final message till time  $t = T^1 - 1$ , it must do so at time  $T^1$ . In other words, we require that  $\tau^1 \leq T^1$ . Similarly for observer 2 described in Case A above, we require that it can at most take  $T^2$  measurements before declaring its final decision, that is,  $\tau^2 \leq T^2$ . The optimization problem is to select policies  $\Gamma^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1}^1)$  and  $\Gamma^2 = (\gamma_0^2, \gamma_1^2, \gamma_2^2, \dots, \gamma_{T^2}^2)$  to minimize

$$\mathbb{E}^{\Gamma^1, \Gamma^2} \{c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \quad (7)$$

where  $\tau^1, \tau^2$  and  $U_k^2, k = 0, 1, \dots$  are defined by equations (2), (3) and (4) above.

2) *Problem P2*: As in Problem P1, we have a finite horizon  $T^1$  for O1, that is,  $\tau^1 \leq T^1$  and a finite horizon  $T^2 (\geq T^1)$  for O2. O2's operation is as described in Case B above. The optimization problem is to select policies  $\Gamma^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1}^1)$  and  $\Gamma^2 = (\gamma_1^2, \gamma_2^2, \dots, \gamma_{T^2}^2)$  to minimize

$$\mathbb{E}^{\Gamma^1, \Gamma^2} \{c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \quad (8)$$

where  $\tau^1, \tau^2$  and  $U_t^2, t = 1, 2, \dots$  are defined by equations (2), (5) and (6) above.

## B. Features of the Problem

In both the problems formulated above, the two observers share a common system objective given by equations (7) or (8). The two observers, however, make decisions based on different information. Thus, Problems P1 and P2 are *team problems*. Moreover, since the actions of observer 1 influence the information available to observer 2, these are dynamic team problems [14]. Dynamic team problems are known to be hard as they usually involve non-convex functional optimization over the space of policies of the decision-makers. Finding structural results for these problems is an important step toward reducing the complexity of these problems. In the next two sections, we establish qualitative properties of the optimal policies of the two observers.

## III. QUALITATIVE PROPERTIES OF OPTIMAL POLICIES FOR OBSERVER O1

### A. Information state for O1

Consider Problem P2 first. We first derive an information state for O1. For that purpose, we define:

$$\pi_t^1 := P(H = 0 | Y_{1:t}^1) \quad (9)$$

The probability  $\pi_t^1$  is observer 1's belief on the hypothesis based on its sequence of observations till time  $t$ . (For  $t = 0$ , we have  $\pi_0^1 = p_0$ ). The following result provides a characterization of O1's optimal policy.

*Theorem 1*: For Problem P2, with an arbitrary but fixed policy  $\Gamma^2$  of O2, there is an optimal policy for O1 of the form:

$$Z_t^1 = \gamma_t^1(\pi_t^1) \quad (10)$$

for  $t = 1, 2, \dots, T^1$ . In particular, if globally optimal policies  $\Gamma^{1,*}, \Gamma^{2,*}$  exist, then  $\Gamma^{1,*}$  can be assumed to be of the form in (10) without loss of optimality. Moreover, for a fixed  $\Gamma^2$ , the optimal policy of O1 can be determined by selecting the minimizing option at each step of the following dynamic program:

$$\begin{aligned} V_{T^1}(\pi) := \min\{ \\ \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | \pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 0], \\ \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | \pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 1] \} \end{aligned} \quad (11)$$

and for  $k = (T^1 - 1), \dots, 2, 1$ ,

$$\begin{aligned} V_k(\pi) := \min\{ \\ \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | \pi_k^1 = \pi, Z_{1:k-1}^1 = b_{1:k-1}, Z_k^1 = 0], \\ \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | \pi_k^1 = \pi, Z_{1:k-1}^1 = b_{1:k-1}, Z_k^1 = 1], \\ c^1 + \mathbb{E}^{\Gamma^2} [V_{k+1}(\pi_{k+1}^1) | \pi_k^1 = \pi, Z_{1:k}^1 = b_{1:k}] \} \end{aligned} \quad (12)$$

where the superscript  $\Gamma^2$  in the expectation denotes that the expectation is defined for a fixed choice of  $\Gamma^2$ . ( $Z_{1:k}^1 = b_{1:k}$  denotes a sequence of  $k$  blank messages.)

*Proof*: See Appendix A. ■

The result of Theorem 1 can be intuitively explained as follows. At any time  $t$ , if the observer 1 has not already sent its final message, it has to choose between three choices of action - send 0, 1 or  $b$ . In order to evaluate the expected cost of sending a 0 or 1, O1 needs a belief on the state of the environment, that is, a belief on  $H$  and a belief on the information available to O2. Since O1 has not yet sent a final message, the information at O2 consists of  $Z_{1:t-1}^1 = b_{1:t-1}$ , the decision of O1 at time  $t$  ( $Z_t^1$ ) and the observations that O2 has made or may make in the future. Thus O1 needs to form a belief on  $Y_{1:T^2}^2$ , since the rest of O2's information it already knows. Now because of conditional independence of observations at the two observers, it is sufficient to form a belief on  $H$  to know the probabilities of  $Y_{1:T^2}^2$ . Similarly, to evaluate the cost of sending a  $b$ , O1 needs to form a belief on O2's information and what information O1 may obtain by future measurements -  $Y_{t+1:T^1}^1$ . Once again, conditional independence of the observations made at different times given  $H$  implies a belief on  $H$  is sufficient to evaluate the cost of this action as well. These arguments indicate that the decisions at O1 should be made based only on its belief on  $H$ , that is,  $\pi_t^1$ .

*Corollary*: Theorem 1 holds for O1 in Problem P1 also.

*Proof:* This result can be obtained by following the steps in Appendix A without any modifications. ■

An intuitive explanation of this result is as follows: In the proof of Theorem 1, we fixed  $\Gamma^2$  to any arbitrary choice. In particular, consider any policy of O2 that waits till it gets a final decision from O1. After it receives the final message from O1 at time  $\tau^1$ , it uses only observations made after  $\tau^1$  to make a decision. This class of policies is essentially the policies available to O2 in problem P1. Since the optimal structure of O1's policy as given in (10) holds for any choice of  $\Gamma^2$ , it also holds for all possible policies of O2 in problem P1.

### B. Classical Two-Threshold Rules Are Not Optimal

In the sequential detection problem with a single observer [13], it is well known that an optimal policy is a function of the observer's belief  $\pi_t$  and is described by two thresholds at each time  $t$ . That is the decision at time  $t$ ,  $Z_t$  is given as:

$$Z_t = \begin{cases} 1 & \text{if } \pi_t \leq \alpha_t \\ N & \text{if } \alpha_t < \pi_t < \beta_t \\ 0 & \text{if } \pi_t \geq \beta_t \end{cases}$$

where  $N$  denotes a decision to continue taking measurements and  $\alpha_t \leq \beta_t$  are real numbers in  $[0, 1]$ . A similar two-threshold structure of optimal policies was also established for the decentralized Wald problem in [11]. We will show by means of two counterexamples that such a structure is not necessarily optimal for observer 1 in Problem P1. Example 2 is similar to an example demonstrating the sub-optimality of threshold rules in a related decentralized sequential detection problem that appeared in [15].

#### Example 1

Consider the following instance of Problem P1. We have equal priors on  $H$ , that is  $P(H = 0) = P(H = 1) = 1/2$ . O1 has a time horizon of  $T^1 = 2$ . Its observation space is  $\mathcal{Y}^1 = \{1, 2, 3\}$ . The observations at time  $t = 1$  have the following conditional probabilities:

Observation	$P(\cdot H = 0)$	$P(\cdot H = 1)$
1	0	$(1 - p)$
2	$p$	$p$
3	$(1 - p)$	0

and at time  $t = 2$  have the following conditional probabilities:

Observation	$P(\cdot H = 0)$	$P(\cdot H = 1)$
1	0	$(1 - q)$
2	$q$	$q$
3	$(1 - q)$	0

where  $p, q \in [0, 1]$ . Observe that O1's belief on the event  $\{H = 0\}$  (that is,  $\pi_t^1$ ), only takes 3 possible values - 0, 1 and  $1/2$  after any number of measurements. O1 has to send a final message - 0 or 1 - to O2 by time  $T^1 = 2$ . If O1 delays sending its final message to time  $t = 2$ , an additional cost  $c^1$  is incurred. After receiving a message from O1, observer 2 can either declare a decision on the hypothesis or take at most 1 more measurement of its own, that is, we have  $T^2 = 1$ . The measurements of O2 are assumed to be noiseless, so when O2 takes a measurement it knows exactly the value of  $H$ .

However, the measurement comes at a cost of  $c^2$ . We assume that  $J(U, H) = 0$  if  $U = H$ , and that the cost of a mistake ( $U \neq H$ ), is sufficiently high so that unless O2 is certain from O1's messages what the true hypothesis is, it will prefer taking a measurement at a cost  $c^2$  rather than making a guess. At  $p = 0.6$ ,  $c^2 > 3c^1$ , it can be easily verified that the best threshold rule for observer 1 is described as follows:

$$Z_1^1 = \begin{cases} 1 & \text{if } \pi_1^1 = 0 \\ b & \text{if } \pi_1^1 = 1/2 \\ 0 & \text{if } \pi_1^1 = 1 \end{cases}$$

and

$$Z_2^1 = \begin{cases} 1 & \text{if } \pi_1^1 = 0 \\ 0 & \text{if } \pi_1^1 > 0 \end{cases}$$

If observer 2 receives 0 or 1 at time  $t = 1$ , it declares the received message as the final decision on the hypothesis, otherwise it waits for a final message from O1. At  $t = 2$ , if O2 receives 1, it declares 1 as the final decision, otherwise it takes a measurement. Then the expected cost of this policy is given as:  $c^1 + pc^1 + p(1+q)c^2/2$  (since, after the observation at time  $t = 1$ , the system incurs additional cost  $c^1$  with probability  $p$  and a cost  $c^2$  with probability  $p/2 + pq/2$ ).

Now consider the following non-threshold policy for observer 1,

$$Z_1^1 = \begin{cases} 1 & \text{if } \pi_1^1 = 0 \\ 0 & \text{if } \pi_1^1 = 1/2 \\ b & \text{if } \pi_1^1 = 1 \end{cases} \quad (13)$$

and

$$Z_2^1 = \begin{cases} 1 & \text{if } \pi_1^1 = 0 \\ 0 & \text{if } \pi_1^1 > 0 \end{cases} \quad (14)$$

Unlike a classical two-threshold rule, the above rule requires O1 to send a blank symbol at time 1 even though O1 is certain that true  $H$  is 0. If observer 2 receives 0 at time  $t = 1$ , it takes a measurement and incurs a cost  $c^2$ . If O2 receives a 1 at  $t = 1$ , it declares 1 as the final decision. If O2 receives a  $b$  at time  $t = 1$ , it waits for the final message at  $t = 2$  and then declares the received message as its final decision on the hypothesis. Then the expected cost of this policy is given as:  $c^1 + pc^2 + (1 - p)c^1/2$  (since, after the observation at time  $t = 1$ , the system incurs a cost  $c^2$  with probability  $p$  and a cost  $c^1$  with probability  $(1 - p)/2$ ). It is now easily seen that at  $p = 0.6$  and  $c^2 > 3c^1$ , if we choose  $q > 1 - \frac{4c^1}{3c^2}$ , the non-threshold policy outperforms the best threshold policy.

*Discussion of the Example:* 1. While the observation model of observer 1 in the above example is non-stationary, the sub-optimality of the classical two-threshold rules is *not* an artifact of the non-stationarity of the observation model. The above example holds for  $p = 0.6$ , and any choice of  $c^1, c^2, q$  such that  $c^2 > 3c^1$  and  $q > 1 - \frac{4c^1}{3c^2}$ . In particular, if we choose  $c^1 = 1$ ,  $c^2 = 3.2$ , a stationary observation model where  $q = p = 0.6$  will give the same conclusion.

2. The principle behind a threshold rule is to stop and send a message if O1 is certain, otherwise postpone the decision and take another measurement. The additional cost of delay is justified by the likelihood of getting a good measurement in the next time instant. In our example, if O1 gets the observations 1 or 3 at  $t = 1$  and is able to convey to O2

that it is certain about the true hypothesis and what this hypothesis is, then it prevents O2 from taking a measurement thus saving a cost  $c^2$ . The threshold rule achieves this objective by sending 0 for observation 3 and 1 for observation 1. However, in the case when O1 gets measurement 2, it decides to wait for the next observation. By choosing  $q$  sufficiently high, the likelihood of getting a good measurement at  $t = 2$  can be made very low. In this case, the cost of delaying a decision ( $c^1$ ) begins outweighing the expected payoff from a new measurement. The non-threshold rule essentially tries to correct this drawback. If at time  $t = 1$ , O1 gets measurement 2, it stops and sends 0 to O2. At O2, this is interpreted as a message to go and take measurement of its own. Note that the non-threshold rule still ensures that whenever O1 is certain about  $H$ , it is able to send enough information to O2 to prevent it from taking a measurement.

#### Example 2

Consider the same problem as in Example 1 but with O1's observations at  $t = 1$  now given by the following conditional probabilities.

Observation	$P(\cdot H = 0)$	$P(\cdot H = 1)$
1	0	$(1 - p)$
2	$p/3$	$2p/3$
3	$2p/3$	$p/3$
4	$(1 - p)$	0

O1's observations at time  $t = 2$  are just noise and give no new information. The rest of the model is same as in Example 1. Note that the observations are indexed in order of the posterior belief  $\pi_1^1$  they generate, that is,  $P(H = 0|Observation1) < P(H = 0|Observation2)$  and so on. If O1 postpones its final message to time  $t = 2$ , it has to pay an additional cost of  $c^1$ . Observer 2 can make a noiseless measurement at a cost of  $c^2$ . As in Example 1, Observer 2's cost of making a wrong decision is chosen sufficiently high so that unless it is certain from O1's message what the true hypothesis is, O2 will prefer taking a measurement at a cost  $c^2$  than making a guess. It can be shown that for equal priors ( $p_0 = 1/2$ ),  $c^2 > 2c^1$  and  $1/2 < p < 1$ , a non-threshold rule for O1 (given below) performs better than any threshold policy.

- At  $t = 1$ , send 0 if observation 2 occurs and 1 if observation 3 occurs. Send a blank otherwise.
- At  $t = 2$ , send 1 if  $\pi_2^1$  is less than  $1/2$  and 0 otherwise.

The corresponding policy for O2 is as follows:

- At  $t = 1$ , if a 0 or 1 is received, take a measurement, otherwise wait till  $t = 2$ .
- At  $t = 2$ , declare the receive symbol as the final decision.

The cost of the above choice of policies is:  $c^1 + pc^2 + (1-p)c^1$ .

### C. Parametric Characterization of Optimal Policies

An important advantage of the threshold rules in the case of the centralized or the decentralized Wald problem is that it modifies the problem of finding the globally optimal policies from a sequential functional optimization problem to a sequential parametric optimization problem. Even though we have established that a classical two-threshold rule does not hold

for our problem, it is still possible to get a finite parametric characterization of an optimal policy for observer 1. Such a parametric characterization provides significant computational advantage in finding optimal policies, for example by reducing the search space for an optimal policy.

In Theorem 1, we have established that for an arbitrarily fixed choice of O2's policy, the optimal policy for O1 can be determined by backward induction using the functions  $V_k(\pi), k = T^1, \dots, 2, 1$ . We will call  $V_k$  the value function at time  $k$ . We have the following lemma.

*Lemma 1:* In problem P1 or P2, with a fixed (but arbitrary) choice of  $\Gamma^2$ , the value function at  $T^1$  can be expressed as:

$$V_{T^1}(\pi) := \min\{L_{T^1}^0(\pi), L_{T^1}^1(\pi)\} \quad (15)$$

where  $L_{T^1}^0(\cdot)$  and  $L_{T^1}^1(\cdot)$  are affine functions of  $\pi$  that depend on the choice of O2's policy  $\Gamma^2$ . Also, the value function at time  $k < T^1$  can be expressed as:

$$V_k(\pi) := \min\{L_k^0(\pi), L_k^1(\pi), G_k(\pi)\} \quad (16)$$

where  $L_k^0(\cdot)$  and  $L_k^1(\cdot)$  are affine functions of  $\pi$  and  $G_k(\cdot)$  is a concave function of  $\pi$ . The functions  $L_k^0(\cdot)$ ,  $L_k^1(\cdot)$  and  $G_k(\cdot)$  depend on the choice of  $\Gamma^2$ .

*Proof:* See Appendix B. ■

*Theorem 2:* In Problem P1 or P2, for any fixed policy  $\Gamma^2$  of O2, an optimal policy for O1 can be characterized by *at most 4 thresholds*. In particular, without any loss in performance, one can assume O1's policy to be of the following form:

$$Z_{T^1}^1 = \begin{cases} 1 & \text{if } \pi_{T^1}^1 \leq \alpha_{T^1} \\ 0 & \text{if } \pi_{T^1}^1 > \alpha_{T^1} \end{cases}$$

where  $0 \leq \alpha_{T^1} \leq 1$  and for  $k = 1, 2, \dots, T^1 - 1$ ,

$$Z_k^1 = \begin{cases} b & \text{if } \pi_k^1 < \alpha_k \\ 1 & \text{if } \alpha_k \leq \pi_k^1 \leq \beta_k \\ b & \text{if } \beta_k < \pi_k^1 < \delta_k \\ 0 & \text{if } \delta_k \leq \pi_k^1 \leq \theta_k \\ b & \text{if } \pi_k^1 > \theta_k \end{cases}$$

where  $0 \leq \alpha_k \leq \beta_k \leq \delta_k \leq \theta_k \leq 1$ .

*Proof:* Theorem 2 is an immediate consequence of Lemma 1, since taking minimum of straight lines and concave functions can partition the interval  $[0, 1]$  into at most five regions. The thresholds above essentially signify the boundaries of these regions. For a given  $\Gamma^2$ , it is possible that at some time instant  $k$ , the optimal policy for O1 partitions the belief interval  $[0, 1]$  as  $\{b, 0, b, 1, b\}$  instead of  $\{b, 1, b, 0, b\}$ . In this case, it is easily seen that simply interchanging the roles of 0 and 1 in O1's policy and in  $\Gamma^2$  at time  $k$  would result in the threshold structure of the theorem without loss of performance. ■

It is of course possible that in specific cases, some of these five regions are absent which would correspond to some of the above thresholds having the same value. For example, in the non-threshold rule given in the Example 1 earlier, the rule at  $t = 1$  corresponds to having  $\alpha = 0$  and  $\beta = \delta$  which results in a 3-interval partition of  $[0, 1]$  corresponding the rule given there.

#### IV. QUALITATIVE PROPERTIES FOR OBSERVER O2

##### A. Problem P1

Consider a fixed policy  $\Gamma^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1}^1)$  for O1. Then, after O1 sends its final message, we can define the following probability for O2:

$$\pi_0^2 := P^{\Gamma^1}(H = 0 | Z_{1:\tau^1}^1)$$

This is O2's belief on the true hypothesis after having observed the messages from O1 (that is a sequence of  $\tau^1 - 1$  blanks and a final  $Z_{\tau^1}^1 \in \{0, 1\}$ ). Now, the optimization problem for O2 is the classical centralized Wald problem [13] with the prior probability given by  $\pi_0^2$ . It is well-known that the optimal policy for the Wald problem is a rule of the form:

$$U_k^2 = \begin{cases} 1 & \text{if } \pi_k^2 \leq w_k^1 \\ N & \text{if } w_k^1 < \pi_k^2 < w_k^2 \\ 0 & \text{if } \pi_k^2 \geq w_k^2 \end{cases}$$

where  $\pi_k^2$  is the belief on hypothesis after  $k$  observations,

$$\begin{aligned} \pi_k^2 &:= P^{\Gamma^1}(H = 0 | Z_{1:\tau^1}^1, Y_{1:k}^2) \\ &= \frac{P(Y_{1:k}^2 | H = 0) \cdot \pi_0^2}{P(Y_{1:k}^2 | H = 0) \cdot \pi_0^2 + P(Y_{1:k}^2 | H = 1) \cdot (1 - \pi_0^2)}, \end{aligned}$$

and  $w_k^1 \leq w_k^2$ , for  $k = 0, 1, 2, \dots, T^2 - 1$  and  $w_{T^2}^1 = w_{T^2}^2$  are the optimal thresholds for the Wald problem with horizon  $T^2$ .

##### B. Information State in Problem P2

Consider a fixed policy  $\Gamma^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1}^1)$  for O1. Define the following probability for O2:

$$\pi_t^2 := P^{\Gamma^1}(H = 0 | Y_{1:t}^2, Z_{1:t}^1)$$

$\pi_t^2$  is observer 2's belief on the hypothesis based on its observations till time  $t$  and the messages received from O1 till time  $t$  (where the messages from O1 could be all blanks or some blanks terminated by a 0 or 1). For  $t = 0$ , we have  $\pi_0^2 = p_0$ .

The following theorem shows that  $\pi_t^2$  and  $Z_{1:t}^1$  together form an information state for O2.

*Theorem 3:* In Problem P2, with an arbitrary but fixed policy  $\Gamma^1$  of O1, there is an optimal policy for O2 of the form:

$$U_t^2 = \gamma_t^2(Z_{1:t}^1, \pi_t^2) \quad (17)$$

for  $t = 1, 2, \dots, T^2$ . Moreover, this optimal policy can be determined by the following dynamic program:

$$\begin{aligned} \tilde{V}_{T^2}(z_{1:T^1}^1, \pi) &:= \min\{\mathbb{E}^{\Gamma^1}[J(0, H) | \pi_{T^2}^2 = \pi], \\ &\quad \mathbb{E}^{\Gamma^1}[J(1, H) | \pi_{T^2}^2 = \pi]\} \end{aligned} \quad (18)$$

and for  $k = (T^1 - 1), \dots, 1$ ,

$$\begin{aligned} \tilde{V}_k(z_{1:k}^1, \pi) &:= \min\{ \\ &\mathbb{E}^{\Gamma^1}[J(0, H) | \pi_k^2 = \pi], \\ &\mathbb{E}^{\Gamma^1}[J(1, H) | \pi_k^2 = \pi], \\ &c^2 + \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \pi_{k+1}^2) | \pi_k^2 = \pi, Z_{1:k}^1 = z_{1:k}^1]\} \end{aligned} \quad (19)$$

*Proof:* See Appendix C. ■

Observe that in the last term of (19), which corresponds to the cost of postponing the final decision at time  $k$ , we have  $\pi_k^2$  as well as all messages from O1 in the conditioning variables. It is because of this term that we need the entire sequence of messages as a part of the information state. To intuitively see why these messages are needed in the conditioning, note that the cost of continuing depends on future messages from O1. In order to form a belief on future messages, O2 needs a belief on the hypothesis, a belief on future observations of O1 and (since O1 has perfect recall) a belief on all observations of O1 so far. Clearly, the messages received till time  $k$  provide information about the observations of O1 till time  $k$  and are therefore included in the information state.

We can now prove the following lemma about the value functions  $\tilde{V}_k$ .

*Lemma 2:* The value function at  $T^2$  can be expressed as:

$$\tilde{V}_{T^2}(z_{1:T^1}^1, \pi) := \min\{l^0(\pi), l^1(\pi)\} \quad (20)$$

where  $l^0$  and  $l^1$  are affine functions of  $\pi$  that are independent of the choice of O1's policy  $\Gamma^1$ . Also, the value function at time  $k$  can be expressed as:

$$\tilde{V}_k(z_{1:k}^1, \pi) := \min\{l^0(\pi), l^1(\pi), G_k(z_{1:k}^1, \pi)\} \quad (21)$$

where, for each realization  $z_{1:k}^1$  of messages from O1,  $G_k$  is a concave function of  $\pi$  that depends on the choice of O1's policy,  $\Gamma^1$ .

*Proof:* See Appendix D. ■

*Theorem 4:* For a fixed policy  $\Gamma^1$  of O1, an optimal policy of O2 is of the form:

$$U_{T^2}^2 = \begin{cases} 1 & \text{if } \pi_{T^2}^2 \leq \alpha_{T^2} \\ 0 & \text{if } \pi_{T^2}^2 > \alpha_{T^2} \end{cases}$$

$$U_k^2 = \begin{cases} 1 & \text{if } \pi_k^2 \leq \alpha_k(Z_{1:k}^1) \\ N & \text{if } \alpha_k(Z_{1:k}^1) < \pi_k^2 < \beta_k(Z_{1:k}^1) \\ 0 & \text{if } \pi_k^2 \geq \beta_k(Z_{1:k}^1) \end{cases}$$

where  $0 \leq \alpha_k(Z_{1:k}^1) \leq \beta_k(Z_{1:k}^1) \leq 1$  are thresholds that depend on sequence of messages received from O1 ( $Z_{1:k}^1$ ).

*Proof:* At any time  $k$ , if  $\pi_k^2 = 0$ , then it is optimal to stop and declare the hypothesis to be 1 since cost of continuing will be at least  $c^2 + J(1, 1)$  which is more than  $J(1, 1)$  - the cost of immediately stopping and declaring  $U_k^2 = 1$ . Similarly, at  $\pi_k^2 = 1$ , it is optimal to stop and declare  $U_k^2 = 0$ . These observations along with the fact that the value functions  $\tilde{V}_k$  are minimum of affine and concave functions for each realization of the messages received imply the result of the theorem. ■

Thus, according to Theorem 4, the thresholds to be used at time  $k$  by O2 depend on the sequence of messages received from O1 until time  $k$ . This kind of parametric characterization may not appear very appealing since for each time  $k$  one may have to know a number of possible thresholds - one for each possible realization of messages  $z_{1:k}^1$ . We will now argue that there is in fact a simple representation of the thresholds. Note that after time  $\tau^1$ , when O1 sends a final message, O2 is faced with a classical Wald problem with an available time-horizon of  $T^2 - \tau^1$ . Now suppose that the classical Wald thresholds are available for a time horizon of length  $T^2$  - lets call these  $(w_0^1, w_0^2), (w_1^1, w_1^2), (w_2^1, w_2^2), \dots, w_{T^2}$ . Then the

Wald thresholds for a problem with time horizon  $T^2 - \tau^1$  are simply  $(w_{\tau^1}^1, w_{\tau^1}^2), (w_{\tau^1+1}^1, w_{\tau^1+1}^2), (w_{\tau^1+2}^1, w_{\tau^1+2}^2), \dots, w_{T^2}$ . Thus, once O2 hears a final message from O1, it starts using the classical Wald thresholds from that time onwards. In other words, O2 operation is described by the following simple algorithm:

- From time  $k = 1$  onwards, the optimal policy is to use a threshold rule given by 2 numbers  $\alpha_k(b_{1:k})$  and  $\beta_k(b_{1:k})$ , until O1 sends its final message  $Z_k^1 \in \{0, 1\}$ . (As before,  $b_{1:k}$  stands for sequence of  $k$  blank messages.)
- If O1 sends the final message at time  $k$ , start using Wald thresholds:  $(w_k^1, w_k^2), \dots, w_{T^2}$ .

Thus O2's optimal policy is completely characterized by just two tables of thresholds:  $[(\alpha_1(b_{1:1}), \beta_1(b_{1:1})), (\alpha_2(b_{1:2}), \beta_2(b_{1:2})), \dots, (\alpha_{T^1}(b_{1:T^1}), \beta_{T^1}(b_{1:T^1}))]$  and the Wald thresholds  $[(w_0^1, w_0^2), (w_1^1, w_1^2), (w_2^1, w_2^2), \dots, w_{T^2}]$ .

## V. OPTIMAL POLICIES

In the previous sections, we identified qualitative properties of the optimal policies for the two observers. Moreover, if the policy  $\Gamma^2$  ( $\Gamma^1$ ) of O2 (O1) has been chosen already, Theorem 1 (Theorem 3) provides a dynamic programming solution to find an optimal policy  $\tilde{\Gamma}^1$  of O1 ( $\tilde{\Gamma}^2$  of O2) for the given choice of  $\Gamma^2$  ( $\Gamma^1$ ). An iterative application of such an approach may be used to identify person-by-person optimal pair of strategies. However, finding globally optimal or near optimal strategies for such dynamic team problems remains a challenging task since it involves non-convex functional optimization [14]. In this section, we will give a sequential decomposition of the global optimization problem. Such a decomposition provides a systematic methodology to find globally optimal or near-optimal policies for the two observers.

### A. Sequential Decomposition for Problem P1

In Problem P1, observer 2 waits to receive a final message from observer 1 before it starts taking its measurements. After receiving the final message, observer 2 is faced with the centralized sequential detection problem studied by Wald. For the Wald problem, the thresholds characterizing the optimal policy and the cost of the optimal policy are known. For a Wald problem with horizon  $T$  and a starting belief  $\pi$  on the event  $\{H = 0\}$ , the cost of using the optimal Wald thresholds is a function of the belief  $\pi$  which we will denote by  $K^T(\pi)$ . Since the Wald thresholds for observer 2 are known (or can be calculated as in [13]), the designer's task in problem P1 is to find the best set of thresholds to be used by observer 1. Finding the best thresholds for all times  $t = 1$  to  $T^1$  is a formidable optimization problem. Firstly, the system objective (equation (7)) is a complicated function of the thresholds selected for observer 1. Moreover, the objective must be optimized over the space of sequences of thresholds to be used from time  $t = 1$  to  $T^1$ . Below, we show that the optimization problem can in principle be solved in a sequential manner. In the resulting sequential decomposition, at each step the optimization is over the set of thresholds to be used at a single time instant instead of the space of sequences of thresholds from time 1 to  $T^1$ . Though the original optimization problem is decomposed

into several "simpler" optimization problems, each of these remain difficult nonetheless. We believe that the decomposed problems may be more amenable to approximation techniques.

We first define the following:

*Definition 1:* Define  $\xi_1 = P(H, \pi_1^1)$  and for  $t = 2, \dots, T^1$  and a given choice of observer 1's decision functions from time instant 1 to  $t - 1$ , that is,  $(\Gamma_{t-1}^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_{t-1}^1))$ , define

$$\xi_t := P^{\Gamma_{t-1}^1}(H, \pi_t^1 | Z_{1:t-1}^1 = b_{1:t-1})$$

For  $t = 1, 2, \dots, T^1$  and for a given choice of functions  $(\Gamma_t^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_t^1))$ , define

$$\eta_t[z_t^1] := P^{\Gamma_t^1}(H, \pi_t^1 | Z_{1:t-1}^1 = b_{1:t-1}, Z_t^1 = z_t^1)$$

where  $z_t^1 \in \{0, 1, b\}$ .

*Lemma 3:* Consider any policy  $\gamma_t^1$ ,  $t = 1, 2, \dots, T^1$  for observer 1 that is characterized by 4 thresholds  $(\alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$ , for  $t = 1, 2, \dots, T^1 - 1$  and a threshold  $\alpha_{T^1}^1$  at time  $T^1$  (as in Theorem 2). Then,

- There exist transformations  $Q_t^1$  for  $t = 1, 2, \dots, T^1$  such that

$$\eta_t[z_t^1] = Q_t^1(\xi_t, \gamma_t^1, z_t^1)$$

for  $z_t^1 \in \{0, 1, b\}$ , and

- There exist transformations  $Q_t^2$ ,  $t = 1, 2, \dots, T^1 - 1$  such that

$$\xi_{t+1} = Q_t^2(\eta_t[b])$$

*Proof:* We first prove the second part of the lemma. By definition,

$$\begin{aligned} \xi_{t+1}(h, \pi^1) &= P^{\Gamma_t^1}(H = h, \pi_{t+1}^1 = \pi^1 | Z_{1:t}^1 = b_{1:t}) \\ &= P^{\Gamma_t^1}(H = h, T_t(\pi_t^1, Y_{t+1}^1) = \pi^1 | Z_{1:t}^1 = b_{1:t}) \end{aligned} \quad (22)$$

where we used the fact that O1's belief at time  $t + 1$  is a function of its belief at time  $t$  and the observation at time  $t + 1$ , that is,  $\pi_{t+1}^1 = T_t(\pi_t^1, Y_{t+1}^1)$  (see Appendix A, equation (48)). The right hand side of (22) can further be written as:

$$\begin{aligned} &= \int_{y, \pi'} [\mathbb{1}_{T_t(\pi', y) = \pi^1} \\ &\quad \cdot P^{\Gamma_t^1}(H = h, \pi_t^1 = \pi', Y_{t+1}^1 = y | Z_{1:t}^1 = b_{1:t})] \end{aligned}$$

where we use  $\int_{y, \pi'}$  as a shorthand for  $\int_{\substack{y \in \mathcal{Y}^1, \\ \pi' \in [0, 1]}}$ . The above integral can be written as:

$$\begin{aligned} &= \int_{y, \pi'} [\mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \\ &\quad \cdot P^{\Gamma_t^1}(H = h, \pi_t^1 = \pi' | Z_{1:t}^1 = b_{1:t})] \\ &= \int_{y, \pi'} [\mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot \eta_t[b](h, \pi')] \end{aligned} \quad (23)$$

The above integral is a function of  $\eta_t[b]$  and known observation statistics. Thus  $\xi_{t+1} = Q_t^2(\eta_t[b])$ , where  $Q_t^2$  is given by the expression in (23).

For the first part of the lemma, consider

$$\begin{aligned} \eta_t[b] &= P^{\Gamma_t^1}(H = h, \pi_t^1 = \pi^1 | Z_{1:t}^1 = b_{1:t}) \\ &= \frac{P^{\Gamma_t^1}(H = h, \pi_t^1 = \pi^1, Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1})}{P^{\Gamma_t^1}(Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1})} \end{aligned} \quad (24)$$

Under the 4-threshold rule for observer 1,  $Z_t^1 = b$  if  $\pi_t^1 \in \mathcal{C}_t$ , where  $\mathcal{C}_t := [0, \alpha_t^1) \cup (\beta_t^1, \delta_t^1) \cup (\theta_t^1, 1]$ . Therefore, the above probability can be written as:

$$\begin{aligned} &= \frac{\mathbb{1}_{\pi_t^1 \in \mathcal{C}_t} \cdot P^{\Gamma_t^1}(H = h, \pi_t^1 = \pi^1 | Z_{1:t-1}^1 = b_{1:t-1})}{P^{\Gamma_t^1}(\pi_t^1 \in \mathcal{C}_t | Z_{1:t-1}^1 = b_{1:t-1})} \\ &= \frac{\mathbb{1}_{\pi_t^1 \in \mathcal{C}_t} \cdot \xi_t(h, \pi^1)}{\sum_{h=0}^1 \int_{\pi'} \mathbb{1}_{\pi' \in \mathcal{C}_t} \xi_t(h, \pi')} \end{aligned} \quad (25)$$

The above equation is a function of  $\xi_t$  and the thresholds selected by  $\gamma_t^1$ . Similar analysis holds for  $\eta_t[0]$  and  $\eta_t[1]$ . This concludes the proof of the lemma. ■

We can now present a sequential decomposition of problem P1.

*Theorem 5:* For  $t = 1, 2, \dots, T^1 - 1$ , there exist functions  $\mathcal{R}_t(\xi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$  and  $\mathcal{R}_t^*(\xi_t)$  where

$$\mathcal{R}_t^*(\xi_t) = \inf_{\alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1} \mathcal{R}_t(\xi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$$

and for  $t = T^1$ , there exist functions  $\mathcal{R}_{T^1}(\xi_{T^1}, \alpha_{T^1}^1)$  and  $\mathcal{R}_{T^1}^*(\xi_{T^1})$  where

$$\mathcal{R}_{T^1}^*(\xi_{T^1}) = \inf_{\alpha_{T^1}^1} \mathcal{R}_{T^1}(\xi_{T^1}, \alpha_{T^1}^1)$$

such that the optimal thresholds can be evaluated from these functions as follows:

- 1) Note that  $\xi_1 := P(H, \pi_1^1)$  is fixed a priori and does not depend on any design choice. The optimal thresholds at  $t = 1$  for O1 are given by optimizing parameters in the definition of  $\mathcal{R}_1^*(\xi_1)$ .
- 2) Once O1's thresholds at  $t = 1$  are fixed,  $\eta_1[b]$  and hence  $\xi_2$  are fixed by lemma 3. The optimal thresholds for O1 at time  $t = 2$  are given by optimizing parameters in the definition of  $\mathcal{R}_2^*(\xi_2)$ .
- 3) Continuing sequentially,  $\xi_t$  is fixed by the choice of past thresholds, and the optimal thresholds for O1 at time  $t$  are given by optimizing parameters in the definition of  $\mathcal{R}_t^*(\xi_t)$ .

*Proof:* We will prove the result by backward induction. Consider first the final horizon for O1:  $T^1$ . Assume that a designer has already specified functions  $\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1-1}^1$  for O1. The designer has to select a function to be used by O1 at time  $T^1$  in case the final message has not been already sent (that is,  $Z_{1:T^1-1}^1 = b_{1:T^1-1}$ ). By Theorem 2, this function is characterized by a single threshold  $\alpha_{T^1}^1$ . For any choice of  $\alpha_{T^1}^1$ , the future cost for the designer is  $K^{T^2}(\pi_0^2)$ , where  $K^{T^2}(\cdot)$  is the cost of using optimal Wald thresholds with a time-horizon  $T^2$  and  $\pi_0^2$  is O2's belief on  $\{H = 0\}$  after receiving  $Z_{1:T^1}^1$ . The expected future cost for the designer can

therefore be expressed as:

$$\begin{aligned} &\mathbb{E}\{c^2\tau^2 + J(U_{\tau^2}, H) | Z_{1:T^1-1}^1 = b_{1:T^1-1}\} \\ &= \mathbb{E}\{K^{T^2}(\pi_0^2) | Z_{1:T^1-1}^1 = b_{1:T^1-1}\} \\ &= K^{T^2}(P(H = 0 | Z_{T^1}^1 = 0, Z_{1:T^1-1}^1 = b_{1:T^1-1})) \\ &\cdot P(Z_{T^1}^1 = 0 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &+ K^{T^2}(P(H = 0 | Z_{T^1}^1 = 1, Z_{1:T^1-1}^1 = b_{1:T^1-1})) \\ &\cdot P(Z_{T^1}^1 = 1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &= K^{T^2}(P(H = 0 | Z_{T^1}^1 = 0, Z_{1:T^1-1}^1 = b_{1:T^1-1})) \\ &\cdot P(\pi_{T^1}^1 > \alpha_{T^1}^1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &+ K^{T^2}(P(H = 0 | Z_{T^1}^1 = 1, Z_{1:T^1-1}^1 = b_{1:T^1-1})) \\ &\cdot P(\pi_{T^1}^1 \leq \alpha_{T^1}^1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &=: L_{T^1}(\eta_{T^1}[0], \eta_{T^1}[1], \xi_{T^1}, \alpha_{T^1}^1) \end{aligned} \quad (26)$$

where we used the fact that the probabilities in the arguments of  $K^{T^2}(\cdot)$  are marginals of  $\eta_{T^1}[0]$   $\eta_{T^1}[1]$  respectively and the probabilities multiplying the functions  $K^{T^2}$  are marginals of  $\xi_{T^1}$ . Using Lemma 3, we can write (26) as

$$\begin{aligned} &L_{T^1}(Q_{T^1}^1(\xi_{T^1}^1, 0, \alpha_{T^1}^1), Q_{T^1}^1(\xi_{T^1}^1, 1, \alpha_{T^1}^1), \xi_{T^1}, \alpha_{T^1}^1) \\ &=: \mathcal{R}_{T^1}(\xi_{T^1}, \alpha_{T^1}^1) \end{aligned}$$

Thus, for a fixed choice of functions  $\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1-1}^1$  used till time  $T^1 - 1$ , the designer's future cost at  $T^1$ , if the final message was not sent before  $T^1$ , is a function of  $\xi_{T^1}$  (that is induced by the choice of the past decision functions) and the threshold  $\alpha_{T^1}^1$  it selects at time  $T^1$ . To find the best choice of threshold, the designer has to select  $\alpha_{T^1}^1$  to minimize  $\mathcal{R}_{T^1}(\xi_{T^1}, \alpha_{T^1}^1)$ . Define

$$\mathcal{R}_{T^1}^*(\xi_{T^1}) = \inf_{\alpha_{T^1}^1} \mathcal{R}_{T^1}(\xi_{T^1}, \alpha_{T^1}^1)$$

For a given  $\xi_{T^1}$ , the function  $\mathcal{R}_{T^1}^*$  describes the optimal future cost for the designer and the optimizing  $\alpha_{T^1}^1$  gives the best threshold.

Now assume that  $\mathcal{R}_{t+1}^*(\xi_{t+1})$  describes the designer's optimal future cost from time  $t + 1$ . At time  $t$ , if the past decision functions  $\gamma_1^1, \gamma_2^1, \dots, \gamma_{t-1}^1$  have been specified already, the designer's task is to select thresholds  $\alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1$  to be used by O1 at time  $t$ . For a given choice of these thresholds, the future cost for the designer is  $K^{T^2}(\pi_0^2)$  if O1 sends a final message at  $t$ . If a blank message is sent at  $t$ , the designer will use the best threshold at the next time  $t + 1$  and the future cost will be  $c^1 + \mathcal{R}_{t+1}^*(\xi_{t+1})$ . The expected future cost for the designer is therefore given as:

$$\begin{aligned} &\mathbb{E}\{c^1(\tau^1 - t) + c^2\tau^2 + J(U_{\tau^2}, H) | Z_{1:t-1}^1 = b_{1:t-1}\} \\ &= K^{T^2}(P(H = 0 | Z_t^1 = 0, Z_{1:t-1}^1 = b_{1:t-1})) \\ &\cdot P(Z_t^1 = 0 | Z_{1:t-1}^1 = b_{1:t-1}) \\ &+ K^{T^2}(P(H = 0 | Z_t^1 = 1, Z_{1:t-1}^1 = b_{1:t-1})) \\ &\cdot P(Z_t^1 = 1 | Z_{1:t-1}^1 = b_{1:t-1}) \\ &+ [c^1 + \mathcal{R}_{t+1}^*(\xi_{t+1})] \cdot P(Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1}) \end{aligned}$$

$$\begin{aligned}
&= K^{T^2} (P(H = 0 | Z_t^1 = 0, Z_{1:t-1}^1 = b_{1:t-1})) \\
&\cdot P(\delta_t^1 < \pi_t^1 < \theta_t^1 | Z_{1:t-1}^1 = b_{1:t-1}) \\
&K^{T^2} (P(H = 0 | Z_t^1 = 1, Z_{1:t-1}^1 = b)) \\
&\cdot P(\alpha_t^1 < \pi_t^1 < \beta_t^1 | Z_{1:t-1}^1 = b_{1:t-1}) \\
&+ [c^1 + \mathcal{R}_{t+1}^*(Q_t^2(\eta_t[b]))] \cdot P(\pi_t^1 \in \mathcal{C}_t | Z_{1:t-1}^1 = b_{1:t-1}) \\
&=: L_t(\eta_t[0], \eta_t[1], \eta_t[b], \xi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1) \quad (27)
\end{aligned}$$

where we used the fact that the probabilities in the arguments of  $K^{T^2}(\cdot)$  are marginals of  $\eta_t[0]$   $\eta_t[1]$  respectively and the probabilities multiplying the functions  $K^{T^2}$  and  $\mathcal{R}_{t+1}^*$  are marginals of  $\xi_t$ . (Recall that  $\mathcal{C}_t := [0, \alpha_t^1) \cup (\beta_t^1, \delta_t^1) \cup (\theta_t^1, 1]$ ). Using Lemma 3, we can write (27) as a function of  $\xi_t$  (that is induced by the choice of past functions used till time  $t - 1$ ) and the thresholds selected at time  $t$ :

$$\mathcal{R}_t(\xi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1) \quad (28)$$

To find the best choice of threshold, the designer has to select  $(\alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$  to minimize  $\mathcal{R}_t(\xi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$ . Define

$$\mathcal{R}_t^*(\xi_t) = \inf_{\substack{\alpha_t^1, \beta_t^1, \\ \delta_t^1, \theta_t^1}} \mathcal{R}_t(\xi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1) \quad (29)$$

For a given  $\xi_t$ , the function  $\mathcal{R}_t^*$  describes the optimal future cost for the designer and the optimizing thresholds are the best thresholds. The above analysis can be inductively repeated for all time instants.

The optimal thresholds can therefore be evaluated as follows: At  $t = 1$ ,  $\xi_1$  is fixed a priori, therefore one can use  $\mathcal{R}_1^*$  to find the best thresholds at time  $t = 1$ . Once these are selected,  $\xi_2$  can be found using Lemma 3 and one can use  $\mathcal{R}_2^*$  to find the best thresholds at time  $t = 2$  and so on. ■

*Discussion:* The problem of choosing the optimal thresholds for observer 1 can be viewed as a sequential problem for the designer as follows: At each time  $t$ , the designer must specify the thresholds to be used by observer 1 in case the final message has not already been sent. In other words, at each time  $t$ , one can think that the designer is aware of the messages sent from O1 to O2 until  $t - 1$  and in case these were only blanks, the designer must choose the thresholds to be used by O1 at time  $t$ . Thus, the designer is faced with a sequential optimization problem with a fixed temporal ordering of its decisions. Observe also that the designer has perfect recall: it knows all messages sent till time  $t$ . The designer, therefore, has a sequential problem with a *classical information structure* [16]. The proof of Theorem 5 essentially describes the dynamic program for the designer's problem. The belief  $\xi_t$  serves as the designer's information state and the functions  $\mathcal{R}_t^*(\xi_t)$  are essentially the value functions of the dynamic program. This approach of introducing a designer with access to the common information between observers (that is, the information known to both observers: the messages from O1 to O2 in Problem P1) so as to convert a decentralized problem to one with classical information structure is illustrated and fully explained in [17, Section IV] for a communication problem. We refer the reader to that paper for a detailed exposition of this approach.

In Problem P1, until the time  $\tau^1$ , the information available to O2 consists only of the messages sent from O1. This

is the same information that the designer uses to select the thresholds. Thus O2 can be thought of as playing the role of the designer in the proof of Theorem 5. The fact that the problem of choosing the thresholds can be viewed from O2's perspective is crucial in determining the nature of the information state for this problem. The form of our information state and the approach of viewing the problem from O2's perspective imitates the information state and the philosophy adopted in [18] for a real-time point-to-point communication problem with noiseless feedback, where the problem of choosing the encoding functions can be viewed from the decoder's perspective.

### B. Sequential Decomposition for Problem P2

In this section, we present a sequential decomposition similar to Theorem 5 for Problem P2. In Problem P2, both observers start taking measurements at time  $t = 1$ . Moreover, O2 is allowed to stop before receiving the final message from O1 (see the time-ordering in Fig.2 for  $t=1,2,\dots$ ). In Problem P2, the messages sent from O1 to O2 are still common information among the two observers. The problem of choosing the optimal thresholds for the two observers can still be viewed as a sequential problem from the perspective of a designer who at any time  $t$  knows the common information. At each time  $t$ , the designer must specify the thresholds to be used by observer 1 in case the final message has not already been sent. It also has to specify -for each realization of messages from O1- the set of thresholds to be used at O2. In other words, at each time  $t$ , one can think that the designer knows the messages sent from O1 to O2 and the designer must choose the thresholds to be used by O1 and O2 at time  $t$ . The designer's problem can therefore be viewed as a sequential optimization problem with classical information structure.

Unlike Problem P1, O2's information no longer coincides with the designer's information of all previous messages from O1, since O2 has its own observations as well. The fact that the designer's problem can no longer be viewed from O2's perspective implies that the information state found for Problem P1 no longer works for this problem. The main challenge now is to find a suitable information state sufficient for performance evaluation for the designer's problem. We present such an information state and the resulting dynamic program below.

As mentioned earlier, once observer 1 has sent its final message to observer 2, the optimization problem for observer 2 becomes the well known centralized sequential detection problem studied by Wald. The thresholds characterizing the optimal policy and the cost of the optimal policy are known. For a Wald problem with horizon  $T$  and a starting belief  $\pi$  on the event  $\{H = 0\}$ , the cost of using the optimal Wald thresholds is a function of the belief  $\pi$  which we denote by  $K^T(\pi)$ . The designer's task is to select the sequence of thresholds to be used by observer 1 and the sequence of thresholds to be used by observer 2 until the final message has been sent from O1 to O2. After O1's final message has been sent, O2's thresholds are known to be the Wald thresholds with appropriate time-horizon. We will now present a sequential decomposition for the designer.

Recall that we defined observer 2's belief on  $H$  as follows:

$$\pi_t^2 := P^{\Gamma^1}(H = 0 | Y_{1:t}^2, Z_{1:t}^1)$$

$\pi_t^2$  evolves in time as O2 gets more measurements and messages. Once O2 has announced its final decision, its belief on  $H$  does not change with time (since O2 is no longer making measurements or listening to messages from O1). We begin with the following definition and lemma.

*Definition 2:* For  $t = 1, 2, \dots, T^1$  and a given choice of observer 1 and observer 2's strategies from time instant 1 to  $t - 1$ , (that is,  $\Gamma_{t-1}^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_{t-1}^1)$  and  $\Gamma_{t-1}^2 = (\gamma_1^2, \gamma_2^2, \dots, \gamma_{t-1}^2)$ ), define

$$D_t := \mathbb{1}_{\tau^2 \geq t}$$

$$\psi_t := P^{\Gamma_{t-1}^1, \Gamma_{t-1}^2}(H, \pi_t^1, \pi_{t-1}^2, D_t | Z_{1:t-1}^1 = b_{1:t-1})$$

where  $\tau^2$  is the stopping time of O2 as defined in (6). For  $t = 1, 2, \dots, T^1 - 1$  and for a given choice of strategies ( $\Gamma_t^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_t^1)$ ) and ( $\Gamma_{t-1}^2 = (\gamma_1^2, \gamma_2^2, \dots, \gamma_{t-1}^2)$ ), define

$$\phi_t[z_t^1] := P^{\Gamma_t^1, \Gamma_{t-1}^2}(H, \pi_t^1, \pi_{t-1}^2, D_t | Z_{1:t-1}^1 = b_{1:t-1}, Z_t^1 = z_t^1)$$

where  $z_t^1 \in \{0, 1, b\}$ .

*Lemma 4:* Consider any policy  $\gamma_t^1$ ,  $t = 1, 2, \dots, T^1$  for observer 1 that is characterized by 4 thresholds  $(\alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$ , for  $t = 1, 2, \dots, T^1 - 1$  and a threshold  $\alpha_{T^1}^1$  at time  $T^1$  (Theorem 2), and a policy  $\gamma_t^2$ ,  $t = 1, 2, \dots, T^2$  which is characterized by thresholds  $(\alpha_t^2, \beta_t^2)$ ,  $t = 1, 2, \dots, T^1 - 1$  to be used if O1 has not sent a final message and the Wald thresholds  $(w_t^1, w_t^2)$ ,  $t = 1, 2, \dots, T^2$  to be used if the final message from O1 has been received. Then, we have:

- i) There exist transformations  $Q_t^1$  for  $t = 1, 2, \dots, T^1$  such that

$$\phi_t[z_t^1] = Q_t^1(\psi_t, \gamma_t^1, z_t^1)$$

for  $z_t^1 \in \{0, 1, b\}$ , and

- ii) There exist transformations  $Q_t^2$ ,  $t = 1, 2, \dots, T^1 - 1$  such that

$$\psi_{t+1} = Q_t^2(\phi_t[b], \gamma_t^2)$$

*Proof:* See Appendix E. ■

We can now present a sequential decomposition of problem P2.

*Theorem 6:* For  $t = 1, 2, \dots, T^1 - 1$ , there exist functions  $\mathcal{F}_t(\psi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$  and  $\mathcal{F}_t^*(\psi_t)$  and  $\mathcal{G}_t(\phi_t[b], \alpha_t^2, \beta_t^2)$  and  $\mathcal{G}_t^*(\phi_t[b])$  where

$$\mathcal{F}_t^*(\psi_t) = \inf_{\alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1} \mathcal{F}_t(\psi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1)$$

$$\mathcal{G}_t^*(\phi_t[b]) = \inf_{\alpha_t^2, \beta_t^2} \mathcal{G}_t(\phi_t[b], \alpha_t^2, \beta_t^2)$$

and for  $t = T^1$ , there exist functions  $\mathcal{F}(\psi_{T^1}, \alpha_{T^1}^1)$  and  $\mathcal{F}_{T^1}^*(\psi_{T^1})$  where

$$\mathcal{F}_{T^1}^*(\psi_{T^1}) = \inf_{\alpha_{T^1}^1} \mathcal{F}(\psi_{T^1}, \alpha_{T^1}^1)$$

such that the optimal thresholds can be evaluated from these functions as follows:

- 1) Note that  $\psi_1$  is fixed a priori and does not depend on any design choice. The optimal thresholds at  $t = 1$  for O1 are given by optimizing parameters in the definition of  $\mathcal{F}_1^*(\psi_1)$ .
- 2) Once O1's thresholds are fixed,  $\phi_1[b]$  is fixed by Lemma 4. The optimizing thresholds to be used by O2 if a blank message was received are given by optimizing parameters in the definition of  $\mathcal{G}_1^*(\phi_1[b])$ . In case a 0 or 1 was received from O1, the optimal thresholds for O2 from this time onwards are the Wald thresholds for a finite horizon  $T^2 - 1$ .
- 3) Continuing sequentially,  $\psi_t$  is fixed by the choice of past thresholds and the optimal thresholds for O1 at time  $t$  are given by optimizing parameters in the definition of  $\mathcal{F}_t^*(\psi_t)$ . Once O1's thresholds are fixed,  $\phi_t[b]$  is fixed by lemma 4. The optimizing thresholds to be used by O2 if a blank message was received are given by optimizing parameters in the definition of  $\mathcal{G}_t^*(\phi_t[b])$ . In case a 0 or 1 was receiver from O1, the optimal thresholds for O2 from this time onwards are the Wald thresholds for a finite horizon  $T^2 - t$ .

*Proof:* See Appendix F. ■

As in Theorem 5, the sequential decomposition in Theorem 6 is a dynamic programming result for the designer's sequential problem of choosing the thresholds for O1 and O2. At time  $t$ ,  $\psi_t$  is the designer's information state just before selecting the four thresholds to be used at O1 to decide its message  $Z_t^1$ , whereas  $\phi_t$  is designer's information state just before selecting the thresholds to be used by O2 to decide  $U_t^2$ . (See Fig. 2). The actual form of the functions  $\mathcal{F}_t$  and  $\mathcal{G}_t$  is obtained by backward induction in Appendix F.

## VI. INFINITE HORIZON PROBLEM

In this section we analyze infinite horizon analogues of problems P1 and P2. We first focus on Problem P2.

### A. Problem P2 with Infinite Horizon

Consider the model of Problem P2 as described in Section II. We assume that the observation model is stationary, that is, the conditional probability of an observation given  $H$ ,  $(P_t^i(\cdot | H))$  does not change with time. We also remove the restriction on the boundedness of the stopping times, that is,  $\tau^1$  and  $\tau^2$  need not be bounded. The optimization problem is to select policies  $\Gamma^1 = (\gamma_1^1, \gamma_2^1, \dots)$  and  $\Gamma^2 = (\gamma_1^2, \gamma_2^2, \dots)$  to minimize

$$\mathbb{E}^{\Gamma^1, \Gamma^2} \{c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \quad (30)$$

where  $\tau^1, \tau^2$  and  $U_t^2$  are defined by equations (2), (5) and (6). We assume that the cost parameters  $c^1, c^2$  are finite positive numbers and that  $J(U^2, H)$  is non-negative and bounded by a constant  $L$  for all  $U^2$  and  $H$ .

*Remark:* We can restrict attention to policies for which  $E\{\tau^1\}$  and  $E\{\tau^2\}$  are finite, since otherwise the expected cost would be infinite. Thus, we have that  $\tau^1$  and  $\tau^2$  are almost surely finite. However, the stopping times may not necessarily be bounded even under optimal policies.

1) *Qualitative Properties for Observer 2:* Consider any fixed policy  $\Gamma^1$  for Observer 1. We will provide structural results on optimal policies for Observer 2 that hold for any choice of  $\Gamma^1$ . Consider the case when observer 2 has not stopped before time  $t$ . Consider a realization  $(y_{1:t}^2, z_{1:t}^1)$  of the information available to O2 at time  $t$  and let  $\bar{\pi}_t^2 = P^{\Gamma^1}(H = 0|y_{1:t}^2, z_{1:t}^1)$  be the realization of O2's belief on  $H$ . Let  $\mathcal{A}^\infty$  be the set of all policies available to O2 at time  $t$  after having observed  $y_{1:t}^2, z_{1:t}^1$ , and let  $\mathcal{A}^{T^2}$  be the subset of policies in  $\mathcal{A}^\infty$  for which the stopping time  $\tau^2$  is less than or equal to a finite horizon  $T^2$ , ( $t \leq T^2 < \infty$ ). Then, from the analysis for the finite-horizon problem P2, we know that there exist value-functions  $\tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$  such that

$$\begin{aligned} & \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2) \\ &= \inf_{\Gamma^2 \in \mathcal{A}^{T^2}} \mathbb{E}^{\Gamma^1} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^2, z_{1:t}^1] \end{aligned} \quad (31)$$

This value-function is the optimal finite horizon cost for observer 2 with horizon  $T^2$ .

We define the following function:

$$\begin{aligned} & \tilde{V}_t^\infty(z_{1:t}^1, y_{1:t}^2) \\ &= \inf_{\Gamma^2 \in \mathcal{A}^\infty} \mathbb{E}^{\Gamma^1} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^2, z_{1:t}^1] \end{aligned} \quad (32)$$

*Lemma 5:* i) The value functions  $\tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$  are non-increasing in  $T^2$  and bounded below by 0, hence the limit  $\lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$  exists.

ii) Moreover,

$$\tilde{V}_t^\infty(z_{1:t}^1, y_{1:t}^2) = \lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$$

*Proof:* See Appendix G. ■

We can now prove the following theorem:

*Theorem 7:* For a fixed policy  $\Gamma^1$  for O1, an optimal policy for O2 is of the form:

$$U_t^2 = \begin{cases} 1 & \text{if } \pi_t^2 \leq \alpha_t(Z_{1:t}^1) \\ N & \text{if } \alpha_t(Z_{1:t}^1) < \pi_t^2 < \beta_t(Z_{1:t}^1) \\ 0 & \text{if } \pi_t^2 \geq \beta_t(Z_{1:t}^1) \end{cases}$$

where  $0 \leq \alpha_t(Z_{1:t}^1) \leq \beta_t(Z_{1:t}^1) \leq 1$  are thresholds that depend on the sequence of messages received from O1 ( $Z_{1:t}^1$ ).

*Proof:* Consider a realization  $y_{1:t}^2, z_{1:t}^1$  of O2's observations and messages from O1. Let  $\bar{\pi}_t^2$  be the realization of O2's belief, where  $\bar{\pi}_t^2 = P^{\Gamma^1}(H = 0|z_{1:t}^1, y_{1:t}^2)$ . Since  $\tilde{V}_t^\infty(z_{1:t}^1, y_{1:t}^2) = \lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$ , it follows that  $V_t^\infty$  is a function only of  $z_{1:t}^1$  and  $\bar{\pi}_t^2$ . Since O2 at time  $t$  has only 3 possible choices, we must have:

$$\begin{aligned} & \tilde{V}_t^\infty(z_{1:t}^1, \bar{\pi}_t^2) := \min\{ \\ & \mathbb{E}^{\Gamma^1} [J(0, H) | \bar{\pi}_t^2], \\ & \mathbb{E}^{\Gamma^1} [J(1, H) | \bar{\pi}_t^2], \\ & c^2 + \mathbb{E}^{\Gamma^1} [\tilde{V}_{t+1}^\infty(Z_{1:t+1}^1, \bar{\pi}_{t+1}^2) | \bar{\pi}_t^2, z_{1:t}^1] \} \end{aligned} \quad (33)$$

From Lemma 2, we know that the first two terms are affine in  $\bar{\pi}_t^2$ . From Lemma 5, we know that  $\tilde{V}_{t+1}^\infty$  is the limit of a sequence of finite-horizon value functions. Now, for a fixed  $z_{1:t+1}^1$ , the finite horizon value functions are concave in  $\bar{\pi}_{t+1}^2$  (from Lemma 2), therefore, for a fixed  $z_{1:t+1}^1$ , the limit  $\tilde{V}_{t+1}^\infty$

is concave in  $\bar{\pi}_{t+1}^2$  as well. Using the concavity of  $\tilde{V}_{t+1}^\infty$  and following the arguments in the proof of Lemma 2, we can show that the third term in equation (33) is concave in  $\bar{\pi}_t^2$  for a fixed  $z_{1:t}^1$ . Thus, for a given realization of  $z_{1:t}^1$ , the infinite horizon value function is minimum of two affine and one concave function. Moreover, it is optimal for O2 to stop if  $\bar{\pi}_t^2 = 0$  or 1. Therefore, the optimal policy for O2 must be of the form:

$$U_t^2 = \begin{cases} 1 & \text{if } \pi_t^2 \leq \alpha_t(Z_{1:t}^1) \\ N & \text{if } \alpha_t(Z_{1:t}^1) < \pi_t^2 < \beta_t(Z_{1:t}^1) \\ 0 & \text{if } \pi_t^2 \geq \beta_t(Z_{1:t}^1) \end{cases}$$

■

As in the finite horizon problem, once observer 1 has sent the final message to observer 2, observer 2 is faced with the classical centralized Wald problem. With an infinite horizon, the optimal Wald policies are characterized by stationary thresholds (say,  $(w^1, w^2)$ ) that do not change with time [13]. Thus, in the infinite horizon version of Problem P2, observer 2's operation can be described by the following algorithm:

- From time  $k = 1$  onwards, the optimal policy is to use a threshold rule given by 2 numbers  $\alpha_k(b_{1:k})$  and  $\beta_k(b_{1:k})$ , until O1 sends its final message  $Z_k^1 \in \{0, 1\}$ .
- From the time O1 sends a final message, start using the stationary Wald thresholds  $(w^1, w^2)$ .

2) *Qualitative Properties for Observer 1:* Consider a fixed policy  $\Gamma^2$  for O2 which belongs to the set of finite horizon policies  $\mathcal{A}^{T^2}$  with horizon  $T^2$ . We will show that given such a policy for O2, Observer 1's infinite horizon optimal policy is characterized by 4 thresholds on its posterior belief. We will employ arguments similar to those used in the previous section.

Consider the case when observer 1 has not stopped before time  $t$ . Consider a realization  $(y_{1:t}^1)$  of the information available to O1 at time  $t$  and let  $\bar{\pi}_t^1 = P(H = 0|y_{1:t}^1)$  be the realization of O1's belief on  $H$ . Let  $\mathcal{B}^\infty$  be the set of all policies available to O2 at time  $t$  after having observed  $y_{1:t}^1$ , and let  $\mathcal{B}^{T^1}$  be the subset of policies in  $\mathcal{B}^\infty$  for which the stopping time  $\tau^1$  is less than or equal to a finite horizon  $T^1$ , ( $t \leq T^1 < \infty$ ). Then, from the analysis for the finite-horizon problem P2, we know that there exist value-functions  $V_t^{T^1}(\bar{\pi}_t^1)$  such that

$$\begin{aligned} & V_t^{T^1}(\bar{\pi}_t^1) \\ &= \inf_{\Gamma^1 \in \mathcal{B}^{T^1}} \mathbb{E}^{\Gamma^2} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \end{aligned} \quad (34)$$

where  $\bar{\pi}_t^1 = P(H = 0|y_{1:t}^1)$ .

We define the following function:

$$\begin{aligned} & V_t^\infty(y_{1:t}^1) \\ &= \inf_{\Gamma^1 \in \mathcal{B}^\infty} \mathbb{E}^{\Gamma^2} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \end{aligned} \quad (35)$$

*Lemma 6:* i) For a fixed finite-horizon policy of O2, the value functions  $V_t^{T^1}(\bar{\pi}_t^1)$  for O1 are non-increasing in  $T^1$  and bounded below by 0, hence the limit  $\lim_{T^1 \rightarrow \infty} V_t^{T^1}(\bar{\pi}_t^1)$  exists.

ii) Moreover,

$$V_t^\infty(y_{1:t}^1) = \lim_{T^1 \rightarrow \infty} V_t^{T^1}(\bar{\pi}_t^1)$$

*Proof:* See Appendix H. ■

We can now state the following theorem:

*Theorem 8:* For a fixed finite-horizon policy  $\Gamma^2$  for O2, an optimal policy for O1 is of the form:

$$Z_t^1 = \begin{cases} b & \text{if } \pi_t^1 < \alpha_t \\ 1 & \text{if } \alpha_t \leq \pi_t^1 \leq \beta_t \\ b & \text{if } \beta_t < \pi_t^1 < \delta_t \\ 0 & \text{if } \delta_t \leq \pi_t^1 \leq \theta_t \\ b & \text{if } \pi_t^1 > \theta_t \end{cases}$$

where  $0 \leq \alpha_t \leq \beta_t \leq \delta_t \leq \theta_t \leq 1$ .

*Proof:* Because of the above Lemma, we conclude that  $V_t^\infty(y_{1:t}^1)$  depends only on the realization  $\bar{\pi}_t^1$  of the belief ( $\bar{\pi}_t^1 = P(H = 0|y_{1:t}^1)$ ). It is, moreover, a concave function of  $\bar{\pi}_t^1$ . The result of the theorem follows by using arguments similar to those in the proof of Lemma 1. ■

*Theorem 9:* There exist globally  $\epsilon$ -optimal policies  $G^1, G^2$  for observers 1 and 2 respectively, such that,  $G^1$  is characterized by 4 time-varying thresholds.

*Proof:* Consider any  $\epsilon/2$ -optimal pair of policies  $\Gamma^1, \Gamma^2$ . Then, by arguments used in Lemma 5, we know that there exist a finite horizon policy  $\Gamma_{T^2}^2$  such that the pair  $\Gamma^1, \Gamma_{T^2}^2$  is at most  $\epsilon/2$  worse than  $\Gamma^1, \Gamma^2$ . Since  $\Gamma_{T^2}^2$  is a finite horizon policy, by theorem 8, we conclude that O1 can use a 4-threshold rule without losing any performance with respect to the policies  $\Gamma^1, \Gamma_{T^2}^2$ . Thus, we have an  $\epsilon$  optimal pair of policies where O1's policy is characterized by 4 time-varying thresholds. ■

## B. Problem P1 with Infinite Horizon

The above analysis for infinite horizon version of Problem P2 can be easily specialized to the case of Problem P1. In particular, observer 2's problem is now the classical Wald problem with infinite horizon; thus its optimal policy is characterized by two stationary thresholds. Moreover, the arguments of Lemma 6 and Theorems 8 and 9 can be repeated without any modification to obtain the same qualitative properties for observer 1 in Problem P1.

## C. Sub-optimality of Time-invariant Policies of O1

In the classical sequential detection problem with a single observer with infinite time horizon and a stationary observation model, it is known that there are time-invariant optimal policies. Are time-invariant policies optimal for observer O1 with infinite time horizon and a stationary observation model? The following example shows that the answer is negative.

### Example 3

Consider the instance of Problem 1 described in Example 1 of Section III-B with infinite time horizon and a stationary observation model for O1 that is same as the model at time  $t = 1$  in Example 1:

Observation	$P(\cdot H=0)$	$P(\cdot H=1)$
1	0	$(1-p)$
2	$p$	$p$
3	$(1-p)$	0

where  $0 < p < 1$ . The observation model of O2 is also stationary and noiseless. As in Example 1, we assume that

$J(U, H) = 0$  if  $U = H$ , and that the cost of a mistake ( $U \neq H$ ), is sufficiently high so that unless O2 is certain from O1's messages what the true hypothesis is, it will prefer taking a measurement at a cost  $c^2$  rather than making a guess. Also, recall that the prior probability of  $H = 0$  is  $1/2$ .

We start by making the following simple observations:

- (i) For all  $t \geq 1$ ,  $\pi_t^1 \in \{0, 1, 1/2\}$ .
- (ii) If  $\pi_t^1 = 0$  (or 1), then  $\pi_{t+1}^1 = 0$  (or 1).
- (iii) The probability that  $\pi_t^1 = 1/2$  is  $p^t$ .

Now consider any stationary rule of the form  $Z_t^1 = \gamma(\pi_t^1)$ , with  $\pi_t^1 \in \{0, 1, 1/2\}$ . We have 3 possible cases:

- 1) If  $\gamma(0) = b$  or  $\gamma(1) = b$ : Suppose  $\gamma(0) = b$ . With probability  $(1-p)/2$ , the posterior at time  $t = 1$  will be 0, which implies that posterior at all subsequent times will be 0. Thus, under the given stationary rule, with probability  $(1-p)/2$ , O1 will never stop and incur infinite costs. Clearly, such rules cannot be optimal. Similar argument holds for  $\gamma(1) = b$ .
- 2) If  $\gamma(1/2) = b$ : We will lower-bound the cost in this case. At any time  $t$ , observer 1 will not stop with probability  $p^t$ . Hence, the total cost is at least  $(c^1 + \sum_{t \geq 1} p^t \cdot c^1) = c^1/1-p$ .
- 3) If  $\gamma(\pi) \neq b$  for any  $\pi$ : In this case, O1 always stops a time  $t = 1$ . It is easy to see that the best rule of this kind is the one for which  $\gamma(0) = 1$  and  $\gamma(1/2) = \gamma(1) = 0$ . The expected cost of this rule is:  $c^1 + c^2/2 + pc^2/2$ .

Now consider the non-stationary rule described by equations (13) and (14) in Example 1. The cost of this rule, as evaluated in Example 1, is  $c^1 + pc^2 + (1-p)c^1/2$ . A little algebra shows that for  $c^2 = 2c^1$  and  $p > 1/\sqrt{3}$ , the non-stationary rule outperforms any stationary rule described above.

## VII. COMMUNICATION WITH M-ARY ALPHABET

Consider models of Problem P1 or P2 with the following modification: when observer 1 chooses to stop taking measurements and send a message to observer 2, it can choose to send one of  $M$  possible choices from the set:  $\{0, 1, \dots, M-1\}$ . Thus, observer 1's message at time  $t$  to observer 2, which is a function of all its observations,

$$Z_t^1 = \gamma_t^1(Y_{1:t}^1), \quad (36)$$

belongs to the set  $\{0, 1, \dots, M-1, b\}$ , where we use  $b$  for blank message, that is, no transmission. The sequence of functions  $\gamma_t^1, t = 1, 2, \dots$ , constitute the *policy* of observer 1. Let  $\tau^1$  be the stopping time when observer sends a final message to observer 2, that is,

$$\tau^1 = \min\{t : Z_t^1 \in \{0, 1, \dots, M-1\}\} \quad (37)$$

Observer 2's operation and the overall system objective are the same as in problem P1 or P2. Then, we have the following result:

*Theorem 10:* In Problems P1 or P2 where observer 1 can send one of  $M$  possible final messages, there is no loss of optimality in restricting attention to policies for observer 1

that are of the form:

$$Z_{T^1}^1 = \begin{cases} M-1 & \text{if } \pi_{T^1}^1 \leq \alpha_{T^1}^{M-1} \\ M-2 & \text{if } \alpha_{T^1}^{M-1} < \pi_{T^1}^1 \leq \alpha_{T^1}^{M-2} \\ \dots & \\ 1 & \text{if } \alpha_{T^1}^2 < \pi_{T^1}^1 \leq \alpha_{T^1}^1 \\ 0 & \text{if } \pi_{T^1}^1 > \alpha_{T^1}^1 \end{cases}$$

where  $0 \leq \alpha_{T^1}^{M-1} \leq \alpha_{T^1}^{M-2} \leq \dots \leq \alpha_{T^1}^1 \leq 1$  are  $M-1$  thresholds and for  $k = 1, 2, \dots, T^1 - 1$ ,

$$Z_k^1 = \begin{cases} b & \text{if } \pi_k^1 < \alpha_k^{M-1} \\ M-1 & \text{if } \alpha_k^{M-1} \leq \pi_k^1 \leq \beta_k^{M-1} \\ b & \text{if } \beta_k^{M-1} < \pi_k^1 < \alpha_k^{M-2} \\ M-2 & \text{if } \alpha_k^{M-2} \leq \pi_k^1 \leq \beta_k^{M-2} \\ \dots & \\ \dots & \\ 1 & \text{if } \alpha_k^1 \leq \pi_k^1 \leq \beta_k^1 \\ b & \text{if } \beta_k^1 < \pi_k^1 < \alpha_k^0 \\ 0 & \text{if } \alpha_k^0 \leq \pi_k^1 \leq \beta_k^0 \\ b & \text{if } \pi_k^1 > \beta_k^0 \end{cases}$$

where  $0 \leq \alpha_k^{M-1} \leq \beta_k^{M-1} \leq \alpha_k^{M-2} \leq \dots \leq \alpha_k^1 \leq \beta_k^1 \leq \alpha_k^0 \leq \beta_k^0 \leq 1$  are  $2M$  thresholds.

*Proof:* It is straightforward to extend the arguments of Theorem 1 to show that for a fixed policy of observer 2 optimal policies of observer 1 are functions of its posterior belief  $\pi_t^1$ . Similarly, the proof of Lemma 1 can be extended to show that the value function for observer 1 is minimum of  $M$  affine functions of the belief, that represent the expected cost of stopping and sending one of the  $M$  symbols, and 1 concave function of the belief that represents the expected cost of continuing. Taking minimum of affine and concave functions will result in  $M$  intervals of the belief space  $[0, 1]$  where it is optimal to stop and send one of the  $M$  symbols. If at some time  $t$ , the symbols are not ordered in the monotonically decreasing way as specified in the result above, one can permute the symbols in policies of O1 and O2 at time  $t$  to get the desired ordering without losing performance. ■

## VIII. EXTENSION TO MULTIPLE SENSORS

In this section, we extend our results to the case when several peripheral sensors similar to observer 1 in Problems P1 and P2 are required to send a single final message to a coordinating sensor (similar to O2) which may be taking its own measurements from  $t = 1$  or may start taking measurements after receiving final messages from all the peripheral sensors. We show that the peripheral sensors have similar parametric characterizations of their optimal policies as observer O1 in Problems P1 and P2. We obtain a characterization of coordinator's strategy that is similar to that of O2.

### A. Problem P2 with Multiple Peripheral Sensors

Consider a group of  $N$  peripheral sensors: S1, S2, ..., SN and a coordinating sensor S0 (see Figure 1). Each sensor can make repeated observations on the random variable  $H$ . As before, we assume that conditioned on  $H$ , the observations at different sensors are independent, and the observations made at different

time instants at any sensor are also independent conditioned on  $H$ .

Each of the peripheral sensors observes its own measurement process  $Y_t^i, i = 1, 2, \dots, N$  and  $t = 1, 2, \dots$ . At any time  $t$ , the  $i^{th}$  peripheral sensor can decide either to stop and send a binary message 0 or 1 to the coordinating sensor or to continue taking measurements. Each time the  $i^{th}$  sensor decides to continue taking measurements, a cost  $c^i$  is incurred. Each peripheral sensor sends only a single final message to the coordinator. The policy  $\Gamma^i := (\gamma_1^i, \gamma_2^i, \dots)$  of  $i^{th}$  sensor is of the form:

$$Z_t^i = \gamma_t^i(Y_{1:t}^i) \quad (38)$$

where  $Z_t^i$  is  $i^{th}$  sensor's message at time  $t$  to the coordinating sensor.  $Z_t^i$  belongs to the set  $\{0, 1, b\}$ , where we use  $b$  for blank message, that is, no transmission. The time  $\tau^i$  is the stopping time when  $i^{th}$  sensor sends a final message to the coordinating sensor, that is,

$$\tau^i = \min\{t : Z_t^i \in \{0, 1\}\} \quad (39)$$

The coordinating sensor observes its own measurement process,  $Y_t^0, t = 1, 2, \dots$ . In addition, it receives messages from all the peripheral sensors (we assume that when the coordinating sensor receives a message it knows which peripheral sensor sent that message). At any time  $t$ , S0 can decide to stop and declare a final decision on the hypothesis or take a new measurement and wait for more messages from the peripheral sensors. Each time S0 postpones its decision on the hypothesis, it incurs a cost  $c^0$ . When S0 announces a final decision  $U$  on the hypothesis, it incurs a cost given as  $J(U, H)$ . Thus, the coordinator's decision at time  $t$  is given as:

$$U_t = \gamma_t^0(Y_{1:t}^0, Z_{1:t}^1, Z_{1:t}^2, \dots, Z_{1:t}^N) \quad (40)$$

$U_t$  belongs to the set  $\{0, 1, N\}$ , where we use  $N$  for a null decision, that is, a decision to continue waiting for more messages and taking more measurements. The sequence of functions  $\Gamma^0 = (\gamma_1^0, \gamma_2^0, \dots)$  is the policy of the coordinating sensor. The time  $\tau^0$  is the stopping time when S0 announces its final decision on the hypothesis, that is,

$$\tau^0 = \min\{t : U_t \in \{0, 1\}\} \quad (41)$$

We consider the following problem.

*Problem P3:* Consider a finite horizon  $T^i$  for the peripheral sensors (that is, we require that  $\tau^i \leq T^i$ ) and a finite horizon  $T^0$  for the coordinating sensor, that is,  $\tau^0 \leq T^0$ . The optimization problem is to select policies  $\Gamma^0, \Gamma^1, \dots, \Gamma^N$  of all the sensors to minimize

$$\mathbb{E}\left\{\sum_{i=0}^N c^i \cdot \tau^i + J(U_{\tau^0}, H)\right\} \quad (42)$$

We now obtain a characterization of the peripheral sensors' optimal policies. For the  $i^{th}$  peripheral sensor, we define

$$\pi_t^i := P(H = 0 | Y_{1:t}^i) \quad (43)$$

*Theorem 11:* For any peripheral sensor  $i$  and any fixed choice of strategies  $\Gamma^j$ , for  $j = 0, 1, \dots, N, j \neq i$ , there is

an optimal policy of the peripheral sensor  $i$  of the form:

$$Z_{T^i}^i = \begin{cases} 1 & \text{if } \pi_{T^i}^i \leq \alpha_{T^i}^i \\ 0 & \text{if } \pi_{T^i}^i > \alpha_{T^i}^i \end{cases}$$

where  $0 \leq \alpha_{T^i}^i \leq 1$ , and for  $k = 1, 2, \dots, T^i - 1$ ,

$$Z_k^i = \begin{cases} b & \text{if } \pi_k^i < \alpha_k^i \\ 1 & \text{if } \alpha_k^i \leq \pi_k^i \leq \beta_k^i \\ b & \text{if } \beta_k^i < \pi_k^i < \delta_k^i \\ 0 & \text{if } \delta_k^i \leq \pi_k^i \leq \theta_k^i \\ b & \text{if } \pi_k^i > \theta_k^i \end{cases}$$

where  $0 \leq \alpha_k^i \leq \beta_k^i \leq \delta_k^i \leq \theta_k^i \leq 1$ .

*Proof:* The main idea of the proof is that once the policies of all sensors except  $i$  are fixed, the optimization problem for the  $i^{\text{th}}$  sensor is similar to the problem for O1 in Problem P2. The  $i^{\text{th}}$  sensor plays the role of O1 in Problem P2 and the coordinating sensor plays the role of O2. The observations of the coordinating sensor at time  $t$  can be defined as:

$$\tilde{Y}_t^0 := (Y_t^0, Z_t^j, j = 1, 2, \dots, N, j \neq i)$$

Note that conditioned on  $H$ , the observations  $\tilde{Y}_t^0$  are independent of the  $i^{\text{th}}$  sensor's observations. We can now follow the arguments of Theorem 1 and 2 to conclude the result for the  $i^{\text{th}}$  peripheral sensor. ■

To find a characterization of the coordinating sensor's policy, we fix the policies of all peripheral sensors and define

$$\pi_t^0 := P(H = 0 | Y_{1:t}^0, Z_{1:t}^1, Z_{1:t}^2, \dots, Z_{1:t}^N) \quad (44)$$

*Theorem 12:* For any fixed choice of policies of the peripheral sensors, the policy of the coordinating sensor is given as

$$U_{T^0} = \begin{cases} 1 & \text{if } \pi_{T^0}^0 \leq \alpha_{T^0} \\ 0 & \text{if } \pi_{T^0}^0 > \alpha_{T^0} \end{cases}$$

$$U_k = \begin{cases} 1 & \text{if } \pi_k^0 \leq \alpha_k(Z_{1:k}^1, Z_{1:k}^2, \dots, Z_{1:k}^N) \\ N & \text{if } \alpha_k(Z_{1:k}^1, Z_{1:k}^2, \dots, Z_{1:k}^N) < \pi_k^0 < \\ & \beta_k(Z_{1:k}^1, Z_{1:k}^2, \dots, Z_{1:k}^N) \\ 0 & \text{if } \pi_k^0 \geq \beta_k(Z_{1:k}^1, Z_{1:k}^2, \dots, Z_{1:k}^N) \end{cases}$$

where  $0 \leq \alpha_k(Z_{1:k}^1, Z_{1:k}^2, \dots, Z_{1:k}^N) \leq \beta_k(Z_{1:k}^1, Z_{1:k}^2, \dots, Z_{1:k}^N) \leq 1$  are thresholds that depend on sequence of messages received from the peripheral sensors.

*Proof:* The proof follows the arguments of Theorem 3 and Theorem 4. ■

### B. Problem P1 with Multiple Peripheral Sensors

In the multiple sensor analogue of Problem P1, the fusion center waits to receive final messages from all sensors before it starts its operation. The qualitative properties of O2 established in Section IV-A hold here as well. That is, after receiving all the messages, the fusion center's problem is a classical Wald problem with the initial belief

$$\pi_0^2 := P(H = 0 | Z_{1:\tau^1}^1, Z_{1:\tau^2}^2, \dots, Z_{1:\tau^N}^N)$$

Moreover, the characterization of the optimal policies of peripheral sensors obtained in Theorem 11 can be established in this case as well by employing essentially the same arguments as in the proof of Theorem 11.

## IX. CONCLUSION

We presented two new models for sequential problems in decentralized detection where peripheral sensors send only their final message to a fusion center/coordinating sensor. In the first model, the fusion center waits to receive the final message from each of the peripheral sensors and then starts taking measurements of its own. In the second model, the coordinating sensor starts taking measurements from the beginning, without waiting for the final messages from the peripheral sensors. We first addressed simple two sensors (observers) version of the problems. We derived structural properties of optimal policies for the two observers with a single, terminal communication from observer 1 to observer 2. We showed that classical two threshold rules no longer hold for observer 1. However, since observer 1's problem is a stopping time problem, a finite parametric characterization of optimal policies is still possible and is described by at most 4 thresholds. We obtained a characterization of observer 2's optimal policy as well. We presented a methodology to find the optimal policies in a sequential manner. We also looked at an infinite horizon version of the problem with stationary observation statistics. We showed that while a 4 parameter characterization of policies is almost optimal for observer 1, these parameters may not be stationary. We then extended the qualitative results to the general scenario with multiple peripheral sensors sending their final messages to a fusion center/coordinating sensor.

### APPENDIX A

#### PROOF OF THEOREM 1

Consider an arbitrary choice  $\Gamma^2 = (\gamma_1^2, \gamma_2^2, \dots, \gamma_{T^2}^2)$  for O2's policy. O2's policy is assumed to be fixed to  $\Gamma^2$  throughout this proof. Note that for a fixed  $\Gamma^2$ ,  $\tau^2$  and  $U_{\tau^2}^2$  are functions of O2's observation sequence  $(Y_1^2, Y_2^2, \dots, Y_{T^2}^2)$  and messages received from O1  $(Z_1^1, \dots, Z_{\tau^1}^1)$ . In other words, a policy of O2 induces a stopping time function  $S^{\Gamma^2}$  and an estimate function  $R^{\Gamma^2}$  defined for all possible realizations of the observations of O2 and messages from O1 such that

$$\tau^2 = S^{\Gamma^2}(Y_{1:T^2}^2, Z_{1:\tau^1}^1) \quad (45)$$

$$U_{\tau^2}^2 = R^{\Gamma^2}(Y_{1:T^2}^2, Z_{1:\tau^1}^1) \quad (46)$$

Also, by a simple application of Bayes' rule, we know that  $\pi_{k+1}^1$  can be updated from  $\pi_k^1$  and  $Y_{k+1}^1$ .

$$\begin{aligned} \pi_{k+1}^1 &= P(H = 0 | Y_{1:k+1}^1) \\ &= \frac{P(Y_{k+1}^1 | H = 0) \pi_k^1}{P(Y_{k+1}^1 | H = 0) \pi_k^1 + P(Y_{k+1}^1 | H = 1) (1 - \pi_k^1)} \end{aligned} \quad (47)$$

Thus, we have that

$$\pi_{k+1}^1 = T_k(\pi_k^1, Y_{k+1}^1) \quad (48)$$

where  $T_k$  is defined by (47).

We will now show that under any policy for O1, the expected future cost at time  $k$  for O1 is lower bounded by the functions  $V_k$  defined in Theorem 1. Consider any policy  $\Gamma^1$  for O1. Under the policies  $\Gamma^1$  and  $\Gamma^2$ , and for a realization

$y_{1:k}^1$  of O1's observations till time  $k$ , let  $W_k(y_{1:k}^1)$  be observer 1's expected future cost at time instant  $k$  if it has not sent its final message before time  $k$ . That is,

$$\begin{aligned} W_k(y_{1:k}^1) &:= \mathbb{E}^{\Gamma^1, \Gamma^2} [c^1 \cdot (\tau^1 - k) + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:k}^1, \\ Z_{1:k-1}^1 &= b_{1:k-1}] \end{aligned} \quad (49)$$

First consider time  $T^1$ . We have

$$\begin{aligned} V_{T^1}(\pi) &:= \min \{ \\ \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | \pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 0] , \\ \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | \pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 1] \} \end{aligned} \quad (50)$$

If observer 1 has not sent a final decision before time  $T^1$ , then under policy  $\Gamma^1$ , O1 will either send 0 or 1 at time  $T^1$ . O1's expected cost to go at  $T^1$ , if it sends a 0 at time  $T^1$  is

$$\begin{aligned} W_{T^1}(y_{1:T^1}^1) &= w(y_{1:T^1}^1, 0) \\ &:= \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:T^1}^1, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 0] \end{aligned} \quad (51)$$

Similarly, if O1 sends a 1 at  $T^1$ , its expected cost to go is

$$\begin{aligned} W_{T^1}(y_{1:T^1}^1) &= w(y_{1:T^1}^1, 1) \\ &:= \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:T^1}^1, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 1] \end{aligned} \quad (52)$$

Consider the expectation in (51). We can write it as

$$\begin{aligned} &\mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:T^1}^1, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 0] \\ &= \mathbb{E}^{\Gamma^2} [c^2 S^{\Gamma^2}(Y_{1:T^2}^2, Z_{1:T^1}^1) + J(R^{\Gamma^2}(Y_{1:T^2}^2, Z_{1:T^1}^1, H) | y_{1:T^1}^1, \\ Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 0)] \quad (53) \\ &= \mathbb{E}^{\Gamma^2} [c^2 S^{\Gamma^2}(Y_{1:T^2}^2, b_{1:T^1-1}, 0) + \\ &J(R^{\Gamma^2}(Y_{1:T^2}^2, b_{1:T^1-1}, 0), H) | y_{1:T^1}^1, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 0] \end{aligned} \quad (54)$$

where we used (45) and (46) in (53) and substituted  $Z_{1:T^1}^1$  in (54) with the values specified in the conditioning term of the expectation. Since the only random variables left in the expectation in (54) are  $Y_{1:T^2}^2$  and  $H$ , we can write this expectation as

$$\begin{aligned} &\sum_{\substack{h=0,1 \\ y_{1:T^2}^2 \in \mathcal{Y}_{1:T^2}^2}} [P(y_{1:T^2}^2, H = h | y_{1:T^1}^1, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ Z_{T^1}^1 = 0) \\ &\times \{c^2 S^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0) + J(R^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0), h)\}] \end{aligned} \quad (55)$$

Consider first the term for  $h = 0$  in (55). Because of the conditional independence of the observations at the two

observers, we can write this term as follows:

$$\begin{aligned} &\sum_{y_{1:T^2}^2 \in \mathcal{Y}_{1:T^2}^2} [P(y_{1:T^2}^2 | H = 0) \cdot P(H = 0 | y_{1:T^1}^1, \\ Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 0) \times \{c^2 S^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0) + \\ &J(R^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0), 0)\}] \\ &= \sum_{y_{1:T^2}^2 \in \mathcal{Y}_{1:T^2}^2} [P(y_{1:T^2}^2 | H = 0) \cdot \pi_{T^1}^1(y_{1:T^1}^1) \\ &\times \{c^2 S^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0) \\ &+ J(R^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0), 0)\}] \end{aligned} \quad (56)$$

Similarly, the term for  $h = 1$  in (55) can be written as,

$$\begin{aligned} &\sum_{y_{1:T^2}^2 \in \mathcal{Y}_{1:T^2}^2} [P(y_{1:T^2}^2 | H = 1) \cdot (1 - \pi_{T^1}^1(y_{1:T^1}^1)) \\ &\times \{c^2 S^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0) + J(R^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0), 1)\}] \end{aligned} \quad (57)$$

Combining equations (56) and (57), we see that the expectation in (55) depends on  $\pi_{T^1}^1(y_{1:T^1}^1)$  and not on the entire sequence  $y_{1:T^1}^1$ . Hence, we can replace  $y_{1:T^1}^1$  by  $\pi_{T^1}^1(y_{1:T^1}^1)$  in the conditioning in (51). Therefore,

$$\begin{aligned} &w_{T^1}(y_{1:T^1}^1, 0) \\ &= \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:T^1}^1, Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 0] \\ &= \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | \pi_{T^1}^1(y_{1:T^1}^1), Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ &Z_{T^1}^1 = 0] \\ &\geq V_{T^1}(\pi_{T^1}^1(y_{1:T^1}^1)) \end{aligned} \quad (58)$$

where we used the definition of  $V_{T^1}$  in (58). Exactly same arguments can be used if O1 sends a 1 at time  $T^1$  to show that

$$\begin{aligned} w_{T^1}(y_{1:T^1}^1, 1) &:= \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:T^1}^1, \\ Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 1] \\ &\geq V_{T^1}(\pi_{T^1}^1(y_{1:T^1}^1)) \end{aligned} \quad (59)$$

Hence, we conclude that the following inequality always holds for policy  $\Gamma^1$ :

$$W_{T^1}(y_{1:T^1}^1) \geq V_{T^1}(\pi_{T^1}^1(y_{1:T^1}^1))$$

Now consider time  $k$ . Assume that

$$W_{k+1}(y_{1:k+1}^1) \geq V_{k+1}(\pi_{k+1}^1(y_{1:k+1}^1))$$

If observer 1 has not sent a final decision before time  $k$ , then it will send either a 0, 1 or  $b$  at time  $k$ . Therefore, O1's expected cost to go at  $k$ ,  $W_k(y_{1:k}^1)$ , is either

$$\begin{aligned} w_k(y_{1:k}^1, 0) &:= \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:k}^1, Z_{1:k-1}^1 = b_{1:k-1}, \\ Z_k^1 = 0] \end{aligned} \quad (60)$$

if  $Z_k^1 = 0$ ; or

$$\begin{aligned} w_k(y_{1:k}^1, 1) &:= \mathbb{E}^{\Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:k}^1, Z_{1:k-1}^1 = b_{1:k-1}, \\ Z_k^1 = 1] \end{aligned} \quad (61)$$

if  $Z_k^1 = 1$ ; or

$$w_k(y_{1:k}^1, b) := c^1 + \mathbb{E}^{\Gamma^2}[W_{k+1}(y_{1:k}^1, Y_{k+1}^1)|y_{1:k}^1, Z_{1:k}^1 = b_{1:k}] \quad (62)$$

if  $Z_k^1 = b$ .

By arguments similar to those used at time  $T^1$ , we can show that  $V_k$  is a lower bound to expressions in (60) and (61). That is,

$$\begin{aligned} & w_k(y_{1:k}^1, z_k^1) \\ &= \mathbb{E}^{\Gamma^2}[c^2\tau^2 + J(U_{\tau^2}^2, H)|y_{1:k}^1, Z_{1:k-1}^1 = b_{1:k-1}, Z_k^1 = z_k^1] \\ &\geq V_k(\pi_k^1(y_{1:k}^1)) \end{aligned} \quad (63)$$

for  $z_k^1 \in \{0, 1\}$ .

Consider equation (62). From the induction hypothesis at time  $k+1$ , we have that

$$W_{k+1}(y_{1:k+1}^1) \geq V_{k+1}(\pi_{k+1}^1(y_{1:k+1}^1))$$

which implies

$$\begin{aligned} & \mathbb{E}^{\Gamma^2}[W_{k+1}(y_{1:k}^1, Y_{k+1}^1)|y_{1:k}^1, Z_{1:k}^1 = b_{1:k}] \\ &\geq \mathbb{E}^{\Gamma^2}[V_{k+1}(\pi_{k+1}^1(y_{1:k}^1, Y_{k+1}^1))|y_{1:k}^1, Z_{1:k}^1 = b_{1:k}] \\ &= \mathbb{E}^{\Gamma^2}[V_{k+1}(T_k(\pi_k^1(y_{1:k}^1), Y_{k+1}^1))|y_{1:k}^1, Z_{1:k}^1 = b_{1:k}] \end{aligned} \quad (64)$$

The above expectation is a function of  $\pi_k^1(y_{1:k}^1)$  and the conditional probability:

$$P(Y_{k+1}^1|y_{1:k}^1, Z_{1:k}^1 = b_{1:k})$$

which can be expressed as:

$$P(Y_{k+1}^1|H=0) \cdot \pi_k^1(y_{1:k}^1) + P(Y_{k+1}^1|H=1) \cdot (1 - \pi_k^1(y_{1:k}^1))$$

Thus the expectation in (64) depends only on  $\pi_k^1(y_{1:k}^1)$  and not the entire sequence  $y_{1:k}^1$ ; Hence, it can be written as:

$$\mathbb{E}^{\Gamma^2}[V_{k+1}(T_k(\pi_k^1(y_{1:k}^1), Y_{k+1}^1))|\pi_k^1(y_{1:k}^1), Z_{1:k}^1 = b]$$

Equations (62) and (64) then imply that

$$\begin{aligned} & w_k(y_{1:k}^1, b) \\ &= c^1 + \mathbb{E}^{\Gamma^2}[W_{k+1}(y_{1:k}^1, Y_{k+1}^1, Z_{k+1}^1)|y_{1:k}^1, Z_{1:k}^1 = b_{1:k}] \\ &\geq c^1 + \mathbb{E}^{\Gamma^2}[V_{k+1}(T_k(\pi_k^1(y_{1:k}^1), Y_{k+1}^1))|\pi_k^1(y_{1:k}^1), Z_{1:k}^1 = b_{1:k}] \\ &= c^1 + \mathbb{E}^{\Gamma^2}[V_{k+1}(\pi_{k+1}^1)|\pi_k^1(y_{1:k}^1), Z_{1:k}^1 = b_{1:k}] \\ &\geq V_k(\pi_k^1(y_{1:k}^1)) \end{aligned} \quad (65)$$

where we used the definition of  $V_k$  in (65). From equations (63) and (65), we conclude that the inequality  $W_k(y_{1:k}^1) \geq V_k(\pi_k^1(y_{1:k}^1))$  is true. Hence, by induction it holds for all  $k = T^1, T^1 - 1, \dots, 2, 1$ . Since  $\Gamma^1$  was arbitrary, we conclude that  $V_k$  are lower bounds on the expected cost to go for O1 under any policy for O1 (with O2's policy fixed at  $\Gamma^2$ ).

A policy  $\Gamma^*$  that always selects the minimizing option in the definition of  $V_k$  for each  $\pi$  will achieve the lower bounds  $V_k$  on  $W_k$  with equality for all  $k$ . Note that the total expected cost of policy  $\Gamma^1$  for O1 is  $c^1 + \mathbb{E}[W_1(Y_1^1)]$  which is greater than  $c^1 + \mathbb{E}[V_1(\pi_1^1(y_1^1))]$  (since we have shown that  $W_1(y_1^1) \geq V_1(\pi_1^1(y_1^1))$ ). Thus, we have that  $\Gamma^*$  also achieves the lower bound on total expected cost for any policy. Hence, it is optimal.

Thus, an optimal policy is given by selecting the minimizing option in the definition of  $V_k$  at each  $\pi$ . This establishes the dynamic program of Theorem 1 and shows that there is an optimal policy of the form:

$$Z_t^1 = \gamma_t^*(\pi_t^1)$$

## APPENDIX B PROOF OF LEMMA 1

Consider the first term in definition of  $V_{T^1}$ . Using functions  $S^{\Gamma^2}$  and  $R^{\Gamma^2}$  from equations (45) and (46) in first term of (50), we get

$$\begin{aligned} & \mathbb{E}^{\Gamma^2}[c^2\tau^2 + J(U_{\tau^2}^2, H)|\pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ & Z_{T^1}^1 = 0] \\ &= \mathbb{E}^{\Gamma^2}[c^2 S^{\Gamma^2}(Y_{1:T^2}^2, Z_{1:T^1}^1) + J(R^{\Gamma^2}(Y_{1:T^2}^2, Z_{1:T^1}^1), H) \\ & |\pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 0] \\ &= \mathbb{E}^{\Gamma^2}[c^2 S^{\Gamma^2}(Y_{1:T^2}^2, b_{1:T^1-1}, 0) + J(R^{\Gamma^2}(Y_{1:T^2}^2, b_{1:T^1-1}, 0), H) \\ & |\pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 0] \end{aligned} \quad (66)$$

where we substituted  $Z_{1:T^1}^1$  in (66) with the values specified in the conditioning term of the expectation. Since the only random variables left in the expectation in (66) are  $Y_{1:T^2}^2$  and  $H$ , we can write this expectation as

$$\begin{aligned} & \sum_{\{h=0,1\}} \sum_{y_{1:T^2}^2 \in \mathcal{Y}_{1:T^2}^2} [ \\ & P(y_{1:T^2}^2, H = h | \pi_{T^1}^1 = \pi, Z_{1:T^1-1}^1 = b, Z_{T^1}^1 = 0) \\ & \{c^2 S^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0) \\ & + J(R^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0), h)\} \end{aligned} \quad (67)$$

Consider first the term for  $h = 0$  in (67). Because of the conditional independence of the observations at the two observers, we can write this term as follows:

$$\begin{aligned} & \sum_{y_{1:T^2}^2 \in \mathcal{Y}_{1:T^2}^2} [P(y_{1:T^2}^2 | H = 0) \cdot \pi \cdot \\ & \{c^2 S^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0) + J(R^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0), 0)\}] \\ &= \pi \times \\ & [ \sum_{y_{1:T^2}^2 \in \mathcal{Y}_{1:T^2}^2} P(y_{1:T^2}^2 | H = 0) \cdot \{c^2 S^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0) \\ & + J(R^{\Gamma^2}(y_{1:T^2}^2, b_{1:T^1-1}, 0), 0)\} \end{aligned} \quad (68)$$

$$= \pi \times A_{T^1}^{\Gamma^2} \quad (69)$$

where  $A_{T^1}^{\Gamma^2}$  is the factor multiplying  $\pi$  in (68). Note that this factor depends only on the choice of O2's policy. Similar arguments for the term corresponding to  $h = 1$  in (67) show that it can be expressed as

$$(1 - \pi) \times B_{T^1}^{\Gamma^2} \quad (70)$$

Equations (69) and (70) imply that first term of (50) is an affine function of  $\pi$ , given as  $A_{T^1}^{\Gamma^2} \cdot \pi + B_{T^1}^{\Gamma^2} \cdot (1 - \pi)$ . Similar arguments hold for the second term of (50). Hence, we have that

$$V_{T^1}(\pi) := \min\{L_{T^1}^0(\pi), L_{T^1}^1(\pi)\}$$

Since  $V_{T^1}$  is minimum of two affine functions, it is a concave function of  $\pi$ .

We now proceed inductively. Assume that  $V_{k+1}$  is a concave function of  $\pi$  and consider  $V_k$ ,

$$\begin{aligned} V_k(\pi) &:= \min\{ \\ &\mathbb{E}^{\Gamma^2} [c^2\tau^2 + J(U_{\tau^2}^2, H)|\pi_k^1 = \pi, Z_{1:k-1}^1 = b_{1:k-1}, Z_k^1 = 0], \\ &\mathbb{E}^{\Gamma^2} [c^2\tau^2 + J(U_{\tau^2}^2, H)|\pi_k^1 = \pi, Z_{1:k-1}^1 = b_{1:k-1}, Z_k^1 = 1], \\ &c^1 + \mathbb{E}[V_{k+1}(T_k(\pi_k^1, Y_{k+1}^1))|\pi_k^1 = \pi, Z_{1:k}^1 = b] \} \end{aligned} \quad (71)$$

Repeating the arguments used for  $V_{T^1}$ , it can be shown that first two terms in (71) are affine functions of  $\pi$ . These are the functions  $L_k^0$  and  $L_k^1$  in Lemma 1. To prove that the third term is concave function of  $\pi$ , we use the induction hypothesis that  $V_{k+1}$  is a concave function of  $\pi$ . Then,  $V_{k+1}$  can be written as an infimum of affine functions

$$V_{k+1}(\pi) = \inf_i \{\lambda_i \pi + \mu_i\} \quad (72)$$

Furthermore, last term in (71) can be written as:

$$\begin{aligned} &c^1 + \mathbb{E}[V_{k+1}(T_k(\pi, Y_{k+1}^1))|\pi_k^1 = \pi, Z_{1:k}^1 = b_{1:k}] \\ &= c^1 + \sum_{y_{k+1}^1 \in \mathcal{Y}^1} [P(y_{k+1}^1|\pi_k^1 = \pi, Z_{1:k}^1 = b_{1:k}) \\ &\quad \cdot V_{k+1}(T_k(\pi, y_{k+1}^1))] \end{aligned} \quad (73)$$

Using the definition of  $T_k$  from equation (47), each term in the above summation can be written as

$$\begin{aligned} &P(y_{k+1}^1|\pi_k^1 = \pi, Z_{1:k}^1 = b_{1:k}) \\ &\cdot V_{k+1} \left( \frac{P(y_{k+1}^1|H=0)\cdot\pi}{P(y_{k+1}^1|H=0)\cdot\pi + P(y_{k+1}^1|H=1)\cdot(1-\pi)} \right) \\ &= \{P(y_{k+1}^1|H=0)\cdot\pi + P(y_{k+1}^1|H=1)\cdot(1-\pi)\} \\ &\cdot V_{k+1} \left( \frac{P(y_{k+1}^1|H=0)\cdot\pi}{P(y_{k+1}^1|H=0)\cdot\pi + P(y_{k+1}^1|H=1)\cdot(1-\pi)} \right) \end{aligned} \quad (74)$$

Now using the characterization of  $V_{k+1}$  in terms of the affine functions (from equation 72) in the equation (74), we obtain

$$\begin{aligned} &\inf_i \{\lambda_i \cdot P(y_{k+1}^1|H=0)\cdot\pi + \\ &(P(y_{k+1}^1|H=0)\cdot\pi + P(y_{k+1}^1|H=1)\cdot(1-\pi)) \cdot \mu_i\} \end{aligned} \quad (75)$$

Substituting this expression in (73), we obtain

$$\begin{aligned} &c^1 + \sum_{y_{k+1}^1 \in \mathcal{Y}^1} [\inf_i \{\lambda_i \cdot P(y_{k+1}^1|H=0)\cdot\pi + \\ &(P(y_{k+1}^1|H=0)\cdot\pi + P(y_{k+1}^1|H=1)\cdot(1-\pi)) \cdot \mu_i\}] \end{aligned} \quad (76)$$

Observe that the expression under the infimum is an affine function of  $\pi$ . Hence, taking the infimum over  $i$  gives a concave function of  $\pi$  for each  $y_{k+1}^1$ . Since the sum of concave functions is concave, the expression in (76) is a concave function of  $\pi$ . We will call this function  $G_k(\pi)$ . Thus, the value function at time  $k$  given by (71) can be expressed as:

$$V_k(\pi) := \min\{L_k^0(\pi), L_k^1(\pi), G_k(\pi)\} \quad (77)$$

Since  $V_k$  is minimum of a concave and two affine functions, it itself is a concave function. This completes the argument for the induction step and (77) now holds for all  $k = (T^1 - 1), \dots, 2, 1$ .

## APPENDIX C PROOF OF THEOREM 3

*Proof:* Let  $\Gamma^1 = (\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1}^1)$  be the fixed policy of O1. By definition of  $\pi_{k+1}^2$ , we have

$$\begin{aligned} \pi_{k+1}^2(Y_{1:k+1}^2, Z_{1:k+1}^1) &:= P^{\Gamma^1}(H=0|Y_{1:k+1}^2, Z_{1:k+1}^1) \\ &= \frac{P(H=0, Y_{k+1}^2, Z_{k+1}^1|Y_{1:k}^2, Z_{1:k}^1)}{\sum_{h=0,1} P(H=h, Y_{k+1}^2, Z_{k+1}^1|Y_{1:k}^2, Z_{1:k}^1)} \end{aligned} \quad (78)$$

(although we omit the superscript  $\Gamma^1$  for ease of notation, it should be understood that these probabilities are defined with a fixed  $\Gamma^1$ .)

Consider the numerator in (78). It can be written as:

$$\begin{aligned} &P(Y_{k+1}^2|H=0, Y_{1:k}^2, Z_{1:k+1}^1) \cdot P(Z_{k+1}^1|H=0, Y_{1:k}^2, Z_{1:k}^1) \cdot \\ &P(H=0|Y_{1:k}^2, Z_{1:k}^1) \\ &= P(Y_{k+1}^2|H=0) \cdot P(Z_{k+1}^1|H=0, Z_{1:k}^1) \cdot \pi_k^2(Y_{1:k}^2, Z_{1:k}^1) \end{aligned} \quad (79)$$

where we used conditional independence of the observations in (79). Under a fixed policy of O1,  $Z_k^1$ s are well-defined random variables and hence the second term in (79) is well-defined. Similar expressions can be obtained for the terms in the denominator of (78). Thus, we have that  $\pi_{k+1}^2$  is a function of  $\pi_k^2$ ,  $Y_{k+1}^2$  and  $Z_{1:k+1}^1$ . That is,

$$\pi_{k+1}^2 = \tilde{T}_k(\pi_k^2, Y_{k+1}^2, Z_{1:k+1}^1) \quad (80)$$

In the statement of Theorem 3, we defined  $\tilde{V}_{T^2}$  as

$$\begin{aligned} \tilde{V}_{T^2}(z_{1:T^1}^1, \pi) &:= \min\{\mathbb{E}^{\Gamma^1}[J(0, H)|\pi_{T^2}^2 = \pi], \\ &\mathbb{E}^{\Gamma^1}[J(1, H)|\pi_{T^2}^2 = \pi]\} \end{aligned} \quad (81)$$

If O2 has not declared a final decision on the hypothesis till  $T^2 - 1$ , and selects  $U_{T^2}^2 = 0$ , then his future cost at time  $T^2$  is

$$\begin{aligned} \tilde{W}_{T^2}(y_{1:T^2}^2, z_{1:T^1}^1, 0) &:= \mathbb{E}^{\Gamma^1}[J(0, H)|y_{1:T^2}^2, z_{1:T^1}^1] \\ &= \pi_{T^2}^2(y_{1:T^2}^2, z_{1:T^1}^1) \cdot J(0, 0) + (1 - \pi_{T^2}^2(y_{1:T^2}^2, z_{1:T^1}^1)) \cdot J(0, 1) \\ &= \mathbb{E}^{\Gamma^1}[J(0, H)|\pi_{T^2}^2(y_{1:T^2}^2, z_{1:T^1}^1)] \end{aligned} \quad (82)$$

which corresponds to the first term in definition of  $\tilde{V}_{T^2}$  at  $\pi_{T^2}^2(y_{1:T^2}^2, z_{1:T^1}^1)$ . A similar expression is true if  $U_{T^2}^2 = 1$ . In either case, we have from the definition of  $\tilde{V}_{T^2}$  that for  $u \in \{0, 1\}$ ,

$$\begin{aligned} \tilde{W}_{T^2}(y_{1:T^2}^2, z_{1:T^1}^1, u) &:= \mathbb{E}^{\Gamma^1}[J(u, H)|y_{1:T^2}^2, z_{1:T^1}^1] \\ &\geq \tilde{V}_{T^2}(z_{1:T^1}^1, \pi_{T^2}^2(y_{1:T^2}^2, z_{1:T^1}^1)) \end{aligned} \quad (83)$$

thus, the optimal action at time  $T^2$  is to select the minimizing option in the definition of  $\tilde{V}_{T^2}$  and the optimal future cost is the value of  $\tilde{V}_{T^2}$ .

We will employ backward induction on the functions  $\tilde{V}_k$  defined in Theorem 3 to show that they represent the optimal value functions for O2. Consider time instant  $k$ . Assume  $\tilde{V}_{k+1}$  gives the optimal cost to go (future cost) function at time  $k+1$ .

We have, by definition,

$$\begin{aligned} \tilde{V}_k(z_{1:k}^1, \pi) &:= \min\{ \\ \mathbb{E}^{\Gamma^1}[J(0, H)|\pi_k^2 = \pi], \\ \mathbb{E}^{\Gamma^1}[J(1, H)|\pi_k^2 = \pi], \\ c^2 + \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \pi_{k+1}^2)|\pi_k^2 = \pi, z_{1:k}^1]\} \end{aligned} \quad (84)$$

At time  $k$ , for a realization  $y_{1:k}^2, z_{1:k}^1$  of O2's observations and O1's messages, the cost of stopping and declaring a decision on the hypothesis at time  $k$  is either

$$\tilde{W}_k(y_{1:k}^2, z_{1:k}^1, 0) := \mathbb{E}^{\Gamma^1}[J(0, H)|y_{1:k}^2, z_{1:k}^1] \quad (85)$$

or

$$\tilde{W}_k(y_{1:k}^2, z_{1:k}^1, 1) := \mathbb{E}^{\Gamma^1}[J(1, H)|y_{1:k}^2, z_{1:k}^1] \quad (86)$$

By arguments similar to those at time  $T^2$ , the above terms are the same as the first two terms of  $\tilde{V}_k(z_{1:k}^1, \pi_k^2(y_{1:k}^2, z_{1:k}^1))$ . The cost of continuing at time  $k$  is

$$\begin{aligned} \tilde{W}_k(y_{1:k}^2, z_{1:k}^1, N) &= c^2 + \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \pi_{k+1}^2)|y_{1:k}^2, z_{1:k}^1] \\ &= c^2 + \end{aligned}$$

$$\begin{aligned} \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \tilde{T}_k(\pi_k^2, Y_{k+1}^2, Z_{1:k+1}^1))|y_{1:k}^2, z_{1:k}^1] \\ = c^2 + \end{aligned} \quad (87)$$

$$\mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(z_{1:k}^1, Z_{k+1}^1, \tilde{T}_k(\pi_k^2, Y_{k+1}^2, z_{1:k}^1, Z_{k+1}^1))|y_{1:k}^2, z_{1:k}^1] \quad (88)$$

The expectation in (88) depends on  $\pi_k^2, z_{1:k}^1$  and  $P^{\Gamma^1}(Y_{k+1}^2, Z_{k+1}^1|y_{1:k}^2, z_{1:k}^1)$ . This probability can be written as:

$$\begin{aligned} P(Y_{k+1}^2|H=0).P(Z_{k+1}^1|H=0, z_{1:k}^1).\pi_k^2 + \\ P(Y_{k+1}^2|H=1).P(Z_{k+1}^1|H=1, z_{1:k}^1).(1 - \pi_k^2) \end{aligned} \quad (89)$$

which depends only on  $z_{1:k}^1$  and  $\pi_k^2$ . Thus, the cost of continuing is the same as

$$c^2 + \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \pi_{k+1}^2)|\pi_k^2(y_{1:k}^2, z_{1:k}^1), z_{1:k}^1]$$

which corresponds to the last term in the definition of  $\tilde{V}_k$ . Consequently, the optimal action at time  $k$  is to select the minimizing option in definition of  $V_k$  and the value of  $V_k$  is the optimal expected cost to go at time  $k$ . This completes the proof of the assertion of Theorem 3. ■

#### APPENDIX D PROOF OF LEMMA 2

*Proof:* The result of Lemma 2 for time  $T^2$  follows from the definition of  $\tilde{V}_{T^2}$  since

$$\mathbb{E}^{\Gamma^1}[J(0, H)|\pi_{T^2}^2 = \pi] = \pi.J(0, 0) + (1 - \pi).J(0, 1)$$

This corresponds to the line  $l^0(\pi)$ . Similarly,

$$\mathbb{E}^{\Gamma^1}[J(1, H)|\pi_{T^2}^2 = \pi] = \pi.J(1, 0) + (1 - \pi).J(1, 1)$$

which corresponds to line  $l^1(\pi)$ . Since, for any realization of  $z_{1:T^2}^1, \tilde{V}_{T^2}$  is minimum of two affine functions of  $\pi$ , it is concave in  $\pi$  for each  $z_{1:T^2}^1$ .

Assume now that  $\tilde{V}_{k+1}(z_{1:k+1}^1, \pi)$  is concave in  $\pi$  for each  $z_{1:k+1}^1$ . Then, we can write  $V_{k+1}$  as:

$$\tilde{V}_{k+1}(z_{1:k+1}^1, \pi) = \inf_i \{\lambda_i(z_{1:k+1}^1).\pi + \mu_i(z_{1:k+1}^1)\} \quad (90)$$

where  $\lambda_i(z_{1:k+1}^1)$  and  $\mu_i(z_{1:k+1}^1)$  are real numbers that depend on  $z_{1:k+1}^1$ . Consider the value-function at time  $k$ .

$$\begin{aligned} \tilde{V}_k(z_{1:k}^1, \pi) &= \min\{\mathbb{E}^{\Gamma^1}[J(0, H)|\pi_k^2 = \pi], \\ &\quad \mathbb{E}^{\Gamma^1}[J(1, H)|\pi_k^2 = \pi], \\ &\quad c^2 + \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \pi_{k+1}^2)|\pi_k^2 = \pi, z_{1:k}^1]\} \end{aligned} \quad (91)$$

The first two terms in (91) correspond to the affine terms  $l^0$  and  $l^1$ . The last term in (91) can be written as:

$$\begin{aligned} c^2 + \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \pi_{k+1}^2)|\pi_k^2 = \pi, z_{1:k}^1] \\ = c^2 + \mathbb{E}^{\Gamma^1}[\tilde{V}_{k+1}(Z_{1:k+1}^1, \tilde{T}_k(\pi_k^2, Y_{k+1}^2, Z_{1:k+1}^1))|\pi_k^2 = \pi, z_{1:k}^1] \\ = c^2 + \sum_{y_{k+1}^2 \in \mathcal{Y}^2} \sum_{z_{k+1}^1 \in \{0,1,b\}} [Pr(y_{k+1}^2, z_{k+1}^1|\pi_k^2 = \pi, z_{1:k}^1). \\ \tilde{V}_{k+1}(z_{1:k+1}^1, \tilde{T}_k(\pi, y_{k+1}^2, z_{1:k+1}^1))] \end{aligned} \quad (92)$$

We now use the fact that  $\tilde{T}_k(\pi, y_{k+1}^2, z_{1:k+1}^1)$  is given as

$$\frac{P(y_{k+1}^2|H=0).P(z_{k+1}^1|H=0, z_{1:k}^1).\pi}{P(y_{k+1}^2, z_{k+1}^1|\pi_k^2 = \pi, z_{1:k}^1)} \quad (93)$$

(see equations (78) and (79)).

Focusing on one term of the summation in (92) and using (90), we can write it as

$$\begin{aligned} P(y_{k+1}^2, z_{k+1}^1|\pi_k^2 = \pi, z_{1:k}^1) \times \\ \inf_i \{\lambda_i(z_{1:k+1}^1) \cdot \left( \frac{P(y_{k+1}^2|H=0).P(z_{k+1}^1|H=0, z_{1:k}^1).\pi}{P(y_{k+1}^2, z_{k+1}^1|\pi_k^2 = \pi, z_{1:k}^1)} \right) \\ + \mu_i(z_{1:k+1}^1)\} \end{aligned} \quad (94)$$

Note that the expression outside the infimum in (94) is the same as the denominator in the term multiplying  $\lambda_i(z_{1:k+1}^1)$  in (94). The expression (94) can now be written as

$$\begin{aligned} \inf_i \{\lambda_i(z_{1:k+1}^1).P(y_{k+1}^2|H=0).P(z_{k+1}^1|H=0, z_{1:k}^1).\pi \\ + \mu_i(z_{1:k+1}^1).P(y_{k+1}^2, z_{k+1}^1|\pi_k^2 = \pi, z_{1:k}^1)\} \end{aligned} \quad (95)$$

Expanding the probability multiplying  $\mu_i$ , we can write (95) as

$$\begin{aligned} \inf_i \{\lambda_i(z_{1:k+1}^1).P(y_{k+1}^2|H=0).P(z_{k+1}^1|H=0, z_{1:k}^1).\pi \\ + \mu_i(z_{1:k+1}^1).(P(y_{k+1}^2|H=0).P(z_{k+1}^1|H=0, z_{1:k}^1).\pi + \\ P(y_{k+1}^2|H=1).P(z_{k+1}^1|H=1, z_{1:k}^1).(1 - \pi))\} \end{aligned} \quad (96)$$

For the given  $z_{1:k+1}^1$  and  $y_{k+1}^2$ , the term in the infimum in (96) is affine in  $\pi$ . Therefore, the expression in (96) is concave in  $\pi$ . Thus, for the given realization of  $z_{1:k}^1$ , each term in the summation in (92) is concave in  $\pi$ . Hence, the sum is concave in  $\pi$  as well. This establishes the structure of  $\tilde{V}_k$  in Lemma 2. To complete the induction argument, we only have to note that since  $\tilde{V}_k$  is the minimum of 2 affine and one concave function of  $\pi$ , it is concave in  $\pi$  (for each  $z_{1:k}^1$ ). ■

APPENDIX E  
PROOF OF LEMMA 4

*Proof:* We first prove the second part of the lemma.

By definition, we have

$$\begin{aligned}
& \psi_{t+1}(h, \pi^1, \pi^2, 1) \\
&= P(H = h, \pi_{t+1}^1 = \pi^1, \pi_t^2 = \pi^2, D_{t+1} = 1 | Z_{1:t}^1 = b_{1:t}) \\
&= P(H = h, T_t(\pi_t^1, Y_{t+1}^1) = \pi^1, \pi_t^2 = \pi^2, \\
&D_{t+1} = 1 | Z_{1:t}^1 = b_{1:t}) \tag{97}
\end{aligned}$$

where we used the fact that OI's belief at time  $t + 1$  is a function of its belief at time  $t$  and the observation at time  $t + 1$ , that is,  $\pi_{t+1}^1 = T_t(\pi_t^1, Y_{t+1}^1)$  (see Appendix A, (48)). The right hand side (RHS) of (97) can further be written as:

$$\begin{aligned}
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(H = h, \pi_t^1 = \pi', Y_{t+1}^1 = y, \pi_t^2 = \pi^2, \\
&D_{t+1} = 1 | Z_{1:t}^1 = b_{1:t}) \\
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot P(H = h, \pi_t^1 = \pi', \\
&\pi_t^2 = \pi^2, D_{t+1} = 1 | Z_{1:t}^1 = b_{1:t}) \\
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot P(H = h, \pi_t^1 = \pi', \\
&\pi_t^2 = \pi^2, D_t \cdot \mathbb{1}_{\alpha_t^2 < \pi_t^2 < \beta_t^2} = 1 | Z_{1:t}^1 = b_{1:t}) \tag{98}
\end{aligned}$$

where we used the fact that if  $Z_{1:t}^1 = b_{1:t}$ , then the event  $\{\tau^2 \geq t + 1\}$  is same as  $\{\tau^2 \geq t\} \cap \{\alpha_t^2 < \pi_t^2 < \beta_t^2\}$  and hence  $D_{t+1} = D_t \cdot \mathbb{1}_{\alpha_t^2 < \pi_t^2 < \beta_t^2}$ . The RHS of (98) can be written as:

$$\begin{aligned}
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot P(H = h, \pi_t^1 = \pi', \\
&\pi_t^2 = \pi^2, D_t = 1 | Z_{1:t}^1 = b_{1:t}) \cdot \mathbb{1}_{\alpha_t^2 < \pi_t^2 < \beta_t^2} \\
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot \phi_t[b](h, \pi', \pi^2, 1) \\
&\cdot \mathbb{1}_{\alpha_t^2 < \pi_t^2 < \beta_t^2} \tag{99}
\end{aligned}$$

The expression given by (99) depends on  $\phi_t[b]$ , the thresholds  $\alpha_t^2, \beta_t^2$  specified by  $\gamma_t^2$  and the observation statistics that are known a priori. Thus  $\psi_{t+1}(h, \pi^1, \pi^2, D_{t+1} = 1)$  is a function of  $\phi_t[b]$  and  $\gamma_t^2$ .

Similarly,  $\psi_{t+1}(h, \pi^1, \pi^2, D_{t+1} = 0)$  can be written as:

$$\begin{aligned}
&\psi_{t+1}(h, \pi^1, \pi^2, D_{t+1} = 0) \\
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot P(H = h, \pi_t^1 = \pi', \\
&\pi_t^2 = \pi^2, D_t \cdot \mathbb{1}_{\alpha_t^2 < \pi_t^2 < \beta_t^2} = 0 | Z_{1:t}^1 = b_{1:t}) \\
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot P(H = h, \pi_t^1 = \pi', \\
&\pi_t^2 = \pi^2, D_t = 0 | Z_{1:t}^1 = b_{1:t}) \\
&+ \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot P(H = h, \pi_t^1 = \pi', \\
&\pi_t^2 = \pi^2, D_t = 1 | Z_{1:t}^1 = b_{1:t}) \cdot \mathbb{1}_{(\pi^2 \leq \alpha_t^2) \cup (\pi^2 \geq \beta_t^2)} \\
&= \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot \phi_t[b](h, \pi', \pi^2, 0) \\
&+ \int_{y, \pi'} \mathbb{1}_{T_t(\pi', y) = \pi^1} \cdot P(Y_{t+1}^1 = y | H = h) \cdot \phi_t[b](h, \pi', \pi^2, 1) \\
&\cdot \mathbb{1}_{(\pi^2 \leq \alpha_t^2) \cup (\pi^2 \geq \beta_t^2)} \tag{100}
\end{aligned}$$

The RHS of (100) depends only on  $\phi_t[b]$  and the thresholds  $\alpha_t^2, \beta_t^2$  specified by  $\gamma_t^2$ . This concludes the proof of the second part of the lemma.

For the first part of the lemma, consider

$$\begin{aligned}
&\phi_t[b](h, \pi^1, \pi^2, 1) \\
&= P(H = h, \pi_t^1 = \pi^1, \pi_t^2 = \pi^2, D_t = 1 | Z_{1:t}^1 = b_{1:t}) \tag{101}
\end{aligned}$$

To simplify this term, first note that

$$\begin{aligned}
&\pi_t^2(y_{1:t}^2, z_{1:t}^1) := P(H = 0 | y_{1:t}^2, z_{1:t}^1) \\
&= \frac{P(H = 0, y_t^2, z_t^1 | y_{1:t-1}^2, z_{1:t-1}^1)}{\sum_{h=0,1} P(H = h, y_t^2, z_t^1 | y_{1:t-1}^2, z_{1:t-1}^1)} \tag{102}
\end{aligned}$$

The numerator in (102) can be written as:

$$\begin{aligned}
&P(y_t^2 | H = 0, y_{1:t-1}^2, z_{1:t}^1) \cdot P(z_t^1 | H = 0, y_{1:t-1}^2, z_{1:t-1}^1) \cdot \\
&P(H = 0 | y_{1:t-1}^2, z_{1:t-1}^1) \\
&= P(y_t^2 | H = 0) \cdot P(z_t^1 | H = 0, z_{1:t-1}^1) \cdot \pi_{t-1}^2(y_{1:t-1}^2, z_{1:t-1}^1) \tag{103}
\end{aligned}$$

where we used the conditional independence of observations given  $H$ . Thus the numerator in (102) can be evaluated from  $y_t^2, \pi_{t-1}^2$  and  $P(z_t^1 | H = 0, z_{1:t-1}^1)$ . Similar expression can be obtained for the terms in the denominator of (102). Therefore, we have

$$\pi_t^2(y_{1:t}^2, z_{1:t}^1) = \tilde{T}_{t-1}(\pi_{t-1}^2, y_t^2, P(z_t^1 | H, z_{1:t-1}^1)) \tag{104}$$

For  $z_{1:t}^1 = b_{1:t}$ , we have

$$\begin{aligned}
&\pi_t^2(y_{1:t}^2, b_{1:t}) = \tilde{T}_{t-1}(\pi_{t-1}^2, y_t^2, P(Z_t^1 = b | H, Z_{1:t-1}^1 = b_{1:t-1})) \\
&= \tilde{T}_{t-1}(\pi_{t-1}^2, y_t^2, P(\pi_t^1 \in \mathcal{C}_t | H, Z_{1:t-1}^1 = b_{1:t-1})) \tag{105}
\end{aligned}$$

where  $\mathcal{C}_t := [0, \alpha_t^1) \cup (\beta_t^1, \delta_t^1) \cup (\theta_t^1, 1]$ . The conditional probability in the argument of  $\tilde{T}_{t-1}$  is a function of  $\psi_t$  and the thresholds specified by  $\gamma_t^1$ . Thus, when  $z_{1:t}^1 = b_{1:t}$ ,

$$\pi_t^2 = \tilde{T}_{t-1}(\pi_{t-1}^2, Y_t^2, \psi_t, \gamma_t^1) \tag{106}$$

(since the function  $\gamma_t^1$  is completely characterized by a set of thresholds, we use  $\gamma_t^1$  to denote the set of thresholds).

Because of (106), the expression in (101) can now be expressed as:

$$\begin{aligned} P(H = h, \pi_t^1 = \pi^1, \tilde{T}_{t-1}(\pi_{t-1}^2, Y_t^2, \psi_t, \gamma_t^1) = \pi^2, \\ D_t = 1 | Z_{1:t}^1 = b_{1:t}) \end{aligned} \quad (107)$$

This can further be expressed as

$$\begin{aligned} \frac{P(H = h, \pi_t^1 = \pi^1, \tilde{T}_{t-1}(\pi_{t-1}^2, Y_t^2, \psi_t, \gamma_t^1) = \pi^2, \\ D_t = 1, Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1})}{P(Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1})} \\ \frac{P(H = h, \pi_t^1 = \pi^1, \tilde{T}_{t-1}(\pi_{t-1}^2, Y_t^2, \psi_t, \gamma_t^1) = \pi^2, \\ D_t = 1, Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1})}{P(\pi_t^1 \in \mathcal{C}_t | Z_{1:t-1}^1 = b_{1:t-1})} \end{aligned} \quad (108)$$

where  $\mathcal{C}_t := [0, \alpha_t^1] \cup (\beta_t^1, \delta_t^1) \cup (\theta_t^1, 1]$ . The denominator is a function of a marginal distribution of  $\psi_t$ . To simplify the numerator, first note that  $\psi_t$  is fixed already by the choice of decision functions till time  $t - 1$ . The numerator in (108) can therefore be written as:

$$\begin{aligned} &= \int_{y, \pi'} \mathbb{1}_{\tilde{T}_{t-1}(\pi', y, \psi_t, \gamma_t^1) = \pi^2} \cdot P(H = h, \pi_t^1 = \pi^1, Y_t^2 = y, \\ &\pi_{t-1}^2 = \pi', D_t = 1, Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1}) \\ &= \int_{y, \pi'} \mathbb{1}_{\tilde{T}_{t-1}(\pi', y, \psi_t, \gamma_t^1) = \pi^2} \cdot P(Y_t^2 = y | H = h) \\ &\cdot P(Z_t^1 = b | \pi_t^1 = \pi^1) \cdot P(H = h, \pi_t^1 = \pi^1, \pi_{t-1}^2 = \pi', \\ &D_t = 1 | Z_{1:t-1}^1 = b) \\ &= \int_{y, \pi'} \mathbb{1}_{\tilde{T}_{t-1}(\pi', y, \psi_t, \gamma_t^1) = \pi^2} \cdot P(Y_t^2 = y | H = h) \cdot \mathbb{1}_{\pi^1 \in \mathcal{C}_t} \cdot \\ &P(H = h, \pi_t^1 = \pi^1, \pi_{t-1}^2 = \pi', D_t = 1 | Z_{1:t-1}^1 = b_{1:t-1}) \\ &= \int_{y, \pi'} \mathbb{1}_{\tilde{T}_{t-1}(\pi', y, \psi_t, \gamma_t^1) = \pi^2} \cdot P(Y_t^2 = y | H = h) \cdot \mathbb{1}_{\pi^1 \in \mathcal{C}_t} \cdot \\ &\psi_t(h, \pi^1, \pi', 1) \end{aligned} \quad (109)$$

The expression in (109) is a function of  $\psi_t$  and the thresholds specified by  $\gamma_t^1$ . Since (109) is equal to the numerator of (108), it follows from (108) and (109), and the fact that the denominator of (108) is a marginal distribution of  $\psi_t$ ,

$$\phi_t[b](h, \pi^1, \pi^2, 1) = Q_t^1(\psi_t, \gamma_t^1, b)$$

Similar analysis holds for  $D_t = 0$  and also for  $\phi_t[0]$  and  $\phi_t[1]$ . ■

#### APPENDIX F PROOF OF THEOREM 6

*Proof:* With the appropriate definitions of the information states  $\psi_t$  and  $\phi_t$ , the proof of Theorem 6 is similar to that of Theorem 5. As in the proof of Theorem 5, we proceed backward in time.

Consider first the final horizon for O1:  $T^1$ . Assume that the designer has already specified functions  $\gamma_1^1, \gamma_2^1, \dots, \gamma_{T^1-1}^1$  for O1 and  $\gamma_1^2, \gamma_2^2, \dots, \gamma_{T^1-1}^2$  for O2. The designer has to select a function to be used by O1 at time  $T^1$  in case O1's final message has not been already sent. By Theorem 2, this function is characterized by a single threshold  $\alpha_{T^1}^1$ . The

expected future cost for the designer is the cost of a Wald problem with horizon  $T^2 - T^1$ , if observer 2 has not already declared its final decision. Thus, the expected cost for the designer is:

$$\begin{aligned} &\mathbb{E}[\{c^2(\tau^2 - T^1) + J(U_{\tau^2}, H)\} \cdot \mathbb{1}_{\tau^2 \geq T^1} | Z_{1:T^1-1}^1 = b_{1:T^1-1}] \\ &= \mathbb{E}[K^{T^2-T^1}(\pi_{T^1}^2) \cdot \mathbb{1}_{\tau^2 \geq T^1} | Z_{1:T^1-1}^1 = b_{1:T^1-1}] \\ &= \mathbb{E}[K^{T^2-T^1}(\pi_{T^1}^2) \cdot D_{T^1} | Z_{1:T^1-1}^1 = b_{1:T^1-1}] \end{aligned} \quad (110)$$

where  $K^{T^2-T^1}(\cdot)$  is the cost of using the optimal Wald thresholds from  $T^1$  onwards with an available time horizon of  $T^2 - T^1$ . This cost can be expressed as:

$$\begin{aligned} &= \mathbb{E}[K^{T^2-T^1}(\pi_{T^1}^2) \cdot D_{T^1} | Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 1] \\ &\cdot P(Z_{T^1}^1 = 1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &+ \mathbb{E}[K^{T^2-T^1}(\pi_{T^1}^2) \cdot D_{T^1} | Z_{1:T^1-1}^1 = b_{1:T^1-1}, Z_{T^1}^1 = 0] \\ &\cdot P(Z_{T^1}^1 = 0 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &= \int_{\pi^2} [K^{T^2-T^1}(\pi^2) \cdot P(\pi_{T^1}^2 = \pi^2, D_{T^1} = 1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ &Z_{T^1}^1 = 1)] \cdot P(\pi_{T^1}^1 \leq \alpha_{T^1}^1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &+ \int_{\pi^2} [K^{T^2-T^1}(\pi^2) \cdot P(\pi_{T^1}^2 = \pi^2, D_{T^1} = 1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}, \\ &Z_{T^1}^1 = 0)] \cdot P(\pi_{T^1}^1 > \alpha_{T^1}^1 | Z_{1:T^1-1}^1 = b_{1:T^1-1}) \\ &=: \mathcal{L}_{T^1}(\phi_{T^1}[0], \phi_{T^1}[1], \psi_{T^1}, \alpha_{T^1}^1) \end{aligned} \quad (111)$$

where we used the fact that the probabilities in the integrals are marginals of  $\phi_{T^1}[1]$   $\phi_{T^1}[0]$  respectively and the probabilities multiplying the integrals are marginals of  $\psi_{T^1}$ . Using Lemma 4, we can write (111) as

$$\begin{aligned} &= \mathcal{L}_{T^1}(Q_{T^1}^1(\psi_{T^1}^1, \alpha_{T^1}^1, 0), Q_{T^1}^1(\psi_{T^1}^1, \alpha_{T^1}^1, 1), \psi_{T^1}, \alpha_{T^1}^1) \\ &=: \mathcal{F}_{T^1}(\psi_{T^1}, \alpha_{T^1}^1) \end{aligned} \quad (112)$$

Thus the optimization problem for the designer is to select  $\alpha_{T^1}^1$  to minimize  $\mathcal{F}_{T^1}(\psi_{T^1}, \alpha_{T^1}^1)$ . Define

$$\mathcal{F}_{T^1}^*(\psi_{T^1}) = \inf_{\alpha_{T^1}^1} \mathcal{F}_{T^1}(\psi_{T^1}, \alpha_{T^1}^1)$$

For a given  $\psi_{T^1}$ , the function  $\mathcal{F}_{T^1}^*$  describes the optimal future cost for the designer and the optimizing  $\alpha_{T^1}^1$  gives the best threshold.

Proceeding backwards, assume  $\mathcal{F}_{t+1}^*$  describes the designer's future cost from time  $t + 1$ . We now consider the designer's problem of selecting thresholds  $\alpha_t^2, \beta_t^2$  to be used by O2 if it received all blank messages from O1, that is,  $Z_{1:t}^1 = b_{1:t}$ . The cost at time  $t$  is  $J(0, H)$  if observer 2 stops and declares 0,  $J(1, H)$  if observer 2 declares 1. In case, observer 2 does not make a final decision at this point, a cost of  $c^2$  is incurred. The future cost for the designer will be the optimal cost at time  $t + 1$  which is given by  $\mathcal{F}_{t+1}^*(\psi_{t+1})$ . Thus the expected

cost is given as:

$$\begin{aligned}
& \mathbb{E}[c^1(\tau^1 - (t+1)) + \{c^2(\tau^2 - t) + J(U_{\tau^2}, H)\} \cdot D_t \\
& |Z_{1:t}^1 = b_{1:t}] \\
& = \mathbb{E}[\{J(1, H) \cdot \mathbb{1}_{\pi_t^2 \leq \alpha_t^2} + J(0, H) \cdot \mathbb{1}_{\pi_t^2 \geq \beta_t^2} + \\
& c^2 \cdot \mathbb{1}_{\pi_t^2 \in [\alpha_t^2, \beta_t^2]}\} \cdot D_t | Z_{1:t}^1 = b_{1:t}] + \mathcal{F}_{t+1}^*(\psi_{t+1}) \\
& = \mathbb{E}[\{J(1, H) \cdot \mathbb{1}_{\pi_t^2 \leq \alpha_t^2} + J(0, H) \cdot \mathbb{1}_{\pi_t^2 \geq \beta_t^2} + \\
& c^2 \cdot \mathbb{1}_{\pi_t^2 \in [\alpha_t^2, \beta_t^2]}\} \cdot D_t | Z_{1:t}^1 = b_{1:t}] + \mathcal{F}_{t+1}^*(Q_t^2(\phi_t[b], \alpha_t^2, \beta_t^2)) \\
& \quad (113) \\
& =: \mathcal{G}_t(\phi_t[b], \alpha_t^2, \beta_t^2) \quad (114)
\end{aligned}$$

where we used the fact that the expectation in (113) depends on the thresholds  $\alpha_t^2, \beta_t^2$ , and the conditional belief on  $H, D_t$  and  $\pi_t^2$  given  $Z_{1:t}^1 = b_{1:t}$  - which is a marginal of  $\phi_t[b]$ . Thus the optimization problem for the designer is to select  $\alpha_t^2, \beta_t^2$  to minimize  $\mathcal{G}_t(\phi_t[b], \alpha_t^2, \beta_t^2)$ . Define

$$\mathcal{G}_t^*(\phi_t[b]) = \inf_{\alpha_t^2, \beta_t^2} \mathcal{G}_t(\phi_t, \alpha_t^2, \beta_t^2)$$

Now consider the designer's problem of selecting thresholds  $\alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1$  to be used by O1 at time  $t$ . The expected future cost is  $K^{T^2-t}(\pi_t^2)$  if a final message is sent at time  $t$  and if O2 had not already stopped (that is,  $D_t = 1$ ). In case a blank message is sent, the designer will need to choose thresholds at time  $t$  for O2 and the optimal future cost would be given by  $c^1 + \mathcal{G}_t^*(\phi_t[b])$ . The total expected future cost is therefore,

$$\begin{aligned}
& \mathbb{E}[c^1(\tau^1 - t) + \{c^2(\tau^2 - t) + J(U_{\tau^2}, H)\} D_t \\
& |Z_{1:t-1}^1 = b_{1:t-1}] \\
& = \mathbb{E}[K^{T^2-t}(\pi_t^2) \cdot D_t | Z_t^1 = 0, Z_{1:t-1}^1 = b_{1:t-1}] \\
& \cdot P(Z_t^1 = 0 | Z_{1:t-1}^1 = b_{1:t-1}) \\
& + \mathbb{E}[K^{T^2-t}(\pi_t^2) \cdot D_t | Z_t^1 = 1, Z_{1:t-1}^1 = b_{1:t-1}] \\
& \cdot P(Z_t^1 = 1 | Z_{1:t-1}^1 = b_{1:t-1}) \\
& + [c^1 + \mathcal{G}_{t+1}^*(\phi_t[b])] \cdot P(Z_t^1 = b | Z_{1:t-1}^1 = b_{1:t-1}) \\
& = \mathbb{E}[K^{T^2-t}(\pi_t^2) \cdot D_t | Z_t^1 = 0, Z_{1:t-1}^1 = b_{1:t-1}] \\
& \cdot P(\delta_t^1 < \pi_t^1 < \theta_t^1 | Z_{1:t-1}^1 = b_{1:t-1}) \\
& + \mathbb{E}[K^{T^2-t}(\pi_t^2) \cdot D_t | Z_t^1 = 1, Z_{1:t-1}^1 = b_{1:t-1}] \\
& \cdot P(\alpha_t^1 < \pi_t^1 < \beta_t^1 | Z_{1:t-1}^1 = b_{1:t-1}) \\
& + [c^1 + \mathcal{G}_{t+1}^*(\phi_t[b]) \cdot D_t] \cdot P(\pi_t^1 \in \mathcal{C}_t | Z_{1:t-1}^1 = b_{1:t-1}) \\
& \quad (115)
\end{aligned}$$

$$=: \mathcal{L}_t(\phi_t[0], \phi_t[1], \phi_t[b], \psi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1) \quad (116)$$

$$=: \mathcal{F}_t(\psi_t, \alpha_t^1, \beta_t^1, \delta_t^1, \theta_t^1) \quad (117)$$

where, to write (116), we used the fact that the two expectations in (115) are functions of  $\phi_t[0]$  and  $\phi_t[1]$  (this can be established using analysis similar to that leading to (111)) and the probabilities multiplying the three terms are marginals of  $\psi_t$ . Further, since  $\phi_t[0]$  and  $\phi_t[1]$  are functions of  $\psi_t$ , we can write (116) as (117). The analysis for time  $t$  can be inductively repeated for all times. ■

## APPENDIX G PROOF OF LEMMA 5

The first part of the lemma follows directly from the fact that  $\tilde{V}_t^{T^2}$  is defined as infimum over a monotonically increasing sequence of sets  $\mathcal{A}^{T^2}$ .

We will now prove the second part of the lemma.  $\tilde{V}_t^\infty$  is defined as infimum of the objective over the set of policies  $\mathcal{A}^\infty$  which contains  $\mathcal{A}^{T^2}, \forall T^2$ , hence we conclude that

$$\tilde{V}_t^\infty(z_{1:t}^1, y_{1:t}^2) \leq \lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2) \quad (118)$$

Assume that the inequality in (118) is strict. Then, there exists a policy  $G \in \mathcal{A}^\infty$  for observer 2 such that the expected cost under  $G$ ,

$$W_t(G) := \mathbb{E}^{\Gamma^1, G}[c^1\tau^1 + c^2\tau^2 + J(U_{\tau^2}, H) | y_{1:t}^2, z_{1:t}^1],$$

is strictly less than  $\lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$ . Therefore, the policy  $G$  is better than any finite horizon policy. We will now construct a sequence of finite horizon policies  $G_{T^2}, T^2 = t, t+1, t+2, \dots$  such that the expected cost of  $G_{T^2}$  approaches the expected cost of policy  $G$  as  $T^2 \rightarrow \infty$ . This will contradict the fact that  $W_t(G) < \lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$ . Let  $\tau^G$  and  $U_{\tau^G}$  be the stopping time and the decision at the stopping time induced under policy  $G$ . The policy  $G_{T^2}$  is characterized by the stopping time  $\tau'$  and the decision at stopping time  $U_{\tau'}$  it induces as follows:

$$\tau' = \begin{cases} \tau^G & \text{if } \tau^G \leq T^2 \\ T^2 & \text{if } \tau^G > T^2 \end{cases}$$

and

$$U_{\tau'} = \begin{cases} U_{\tau^G} & \text{if } \tau^G \leq T^2 \\ 0 & \text{if } \tau^G > T^2 \end{cases}$$

Note that  $G_{T^2}$  is finite horizon policy since it always stops no later than the horizon  $T^2$ . Define

$$W_t(G_{T^2}) := \mathbb{E}^{\Gamma^1, G_{T^2}}[c^1\tau^1 + c^2\tau^2 + J(U_{\tau^2}, H) | y_{1:t}^2, z_{1:t}^1]$$

By assumption, the cost under policy  $G$  is better than the cost under any finite horizon policy. Therefore,  $W_t(G_{T^2}) \geq W_t(G)$ . Moreover,

$$\begin{aligned}
& W_t(G_{T^2}) - W_t(G) \\
& = \mathbb{E}^{\Gamma^1, G_{T^2}}[c^1\tau^1 + c^2\tau^2 + J(U_{\tau^2}, H) | y_{1:t}^2, z_{1:t}^1] \\
& - \mathbb{E}^{\Gamma^1, G}[c^1\tau^1 + c^2\tau^2 + J(U_{\tau^2}, H) | y_{1:t}^2, z_{1:t}^1] \\
& = \mathbb{E}\{c^2(\tau' - \tau^G) + J(U_{\tau'}, H) - J(U_{\tau^G}, H) | y_{1:t}^2, z_{1:t}^1\} \\
& = \mathbb{E}\{c^2(\tau' - \tau^G) + \\
& J(U_{\tau'}, H) - J(U_{\tau^G}, H)\} \cdot \mathbb{1}_{\tau^G \leq T^2} | y_{1:t}^2, z_{1:t}^1 \\
& + \mathbb{E}\{c^2(\tau' - \tau^G) + \\
& J(U_{\tau'}, H) - J(U_{\tau^G}, H)\} \cdot \mathbb{1}_{\tau^G > T^2} | y_{1:t}^2, z_{1:t}^1 \quad (119)
\end{aligned}$$

The first expectation in equation (119) is 0 since for  $\tau^G \leq T^2$ , the policy  $G_{T^2}$  has the same stopping time and the final

decision as policy  $G$ . Thus, we get:

$$\begin{aligned} & W_t(G_{T^2}) - W_t(G) \\ &= \mathbb{E}[\{c^2(\tau' - \tau^G) + \\ & \quad J(U_{\tau'}, H) - J(U_{\tau^G}, H)\} \cdot \mathbb{1}_{\tau^G > T^2} | y_{1:t}^2, z_{1:t}^1] \\ &= \mathbb{E}[\{c^2(T^2 - \tau^G) \\ & \quad + J(0, H) - J(U_{\tau^G}, H)\} \cdot \mathbb{1}_{\tau^G > T^2} | y_{1:t}^2, z_{1:t}^1] \\ &\leq \mathbb{E}[\{J(0, H) - J(U_{\tau^G}, H)\} \cdot \mathbb{1}_{\tau^G > T^2} | y_{1:t}^2, z_{1:t}^1] \quad (120) \\ &\leq L \cdot \mathbb{E}[\mathbb{1}_{\tau^G > T^2} | y_{1:t}^2, z_{1:t}^1] \\ &= L \cdot P(\tau^G > T^2 | y_{1:t}^2, z_{1:t}^1), \quad (121) \end{aligned}$$

where  $L$  is the finite positive constant that upper-bounds  $J(U, H)$ . Since the stopping time under policy  $G$  is almost surely finite (otherwise cost of policy would be infinite), we have that  $P(\tau^G > T^2 | y_{1:t}^2, z_{1:t}^1) \rightarrow 0$ , as  $T^2 \rightarrow \infty$ . Thus, for any  $\epsilon > 0$ , there exists a horizon  $T^2$  large enough for which  $W_t(G_{T^2}) - W_t(G) \leq \epsilon$ . Therefore,

$$\lim_{T^2 \rightarrow \infty} W_t(G_{T^2}) = W_t(G)$$

Hence, we conclude that there does not exist any policy  $G \in \mathcal{A}^\infty$  for which  $W_t(G) < \lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$ . Therefore,  $\tilde{V}_t^\infty(z_{1:t}^1, y_{1:t}^2) = \lim_{T^2 \rightarrow \infty} \tilde{V}_t^{T^2}(z_{1:t}^1, \bar{\pi}_t^2)$ .

#### APPENDIX H PROOF OF LEMMA 6

The first part of the lemma follows directly from the fact that  $V_t^{T^2}$  is defined as infimum over a monotonically increasing sequence of sets  $\mathcal{B}^{T^2}$ .

We will now prove the second part of the lemma. Since  $\mathcal{B}^\infty$  contains  $\mathcal{B}^{T^1}, \forall T^1$ , we conclude that  $V_t^\infty(y_{1:t}^1) \leq \lim_{T^1 \rightarrow \infty} V_t^{T^1}(\bar{\pi}_t^1)$ . Assume that the inequality is strict. Then, there exists a policy  $\Lambda \in \mathcal{B}^\infty$  for observer 1 such that the expected cost under  $\Lambda$ ,

$$W_t(\Lambda) := \mathbb{E}^{\Lambda, \Gamma^2} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1],$$

is strictly less than  $\lim_{T^1 \rightarrow \infty} V_t^{T^1}(\bar{\pi}_t^1)$ . Therefore, the policy  $\Lambda$  is better than any finite horizon policy. We will now construct a sequence of finite horizon policies  $\Lambda_{T^1}, T^1 = t, t+1, t+2, \dots$  such that the expected cost of  $\Lambda_{T^1}$  approaches the cost of policy  $\Lambda$  as  $T^1 \rightarrow \infty$ . This will contradict the fact that  $W_t(\Lambda) < \lim_{T^1 \rightarrow \infty} V_t^{T^1}(\bar{\pi}_t^1)$ .

Let  $\tau^\Lambda$  and  $Z_{\tau^\Lambda}^1$  be the stopping time and the decision at the stopping time induced under policy  $\Lambda$ . The policy  $\Lambda_{T^1}$  is characterized by the stopping time  $\tau^*$  and the decision at stopping time  $Z_{\tau^*}^1$  it induces as follows:

$$\tau^* = \begin{cases} \tau^\Lambda & \text{if } \tau^\Lambda \leq T^1 \\ T^1 & \text{if } \tau^\Lambda > T^1 \end{cases}$$

and

$$Z_{\tau^*}^1 = \begin{cases} Z_{\tau^\Lambda}^1 & \text{if } \tau^\Lambda \leq T^1 \\ 0 & \text{if } \tau^\Lambda > T^1 \end{cases}$$

Note that  $\Lambda_{T^1}$  is finite horizon policy since it always stops no later than the horizon  $T^1$ . Define

$$W_t(\Lambda_{T^1}) := \mathbb{E}^{\Lambda_{T^1}, \Gamma^2} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1, z_{1:t}^1]$$

By assumption, the cost under policy  $\Lambda$  is better than cost under any finite horizon policy. Therefore,  $W_t(\Lambda_{T^1}) \geq W_t(\Lambda)$ . Moreover,

$$\begin{aligned} & W_t(\Lambda_{T^1}) - W_t(\Lambda) \\ &= \mathbb{E}^{\Lambda_{T^1}, \Gamma^2} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \\ &\quad - \mathbb{E}^{\Lambda, \Gamma^2} [c^1 \tau^1 + c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \\ &= \mathbb{E}[c^1(\tau^* - \tau^\Lambda) | y_{1:t}^1] + \mathbb{E}^{\Lambda_{T^1}, \Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \\ &\quad - \mathbb{E}^{\Lambda, \Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \quad (122) \\ &\leq \mathbb{E}^{\Lambda_{T^1}, \Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \\ &\quad - \mathbb{E}^{\Lambda, \Gamma^2} [c^2 \tau^2 + J(U_{\tau^2}^2, H) | y_{1:t}^1] \quad (123) \end{aligned}$$

where we used the fact that since  $\tau^* \leq \tau^\Lambda$ , the first term in (122) is less than or equal to 0. Further, (123) can be written as:

$$\begin{aligned} & \mathbb{E}^{\Lambda_{T^1}, \Gamma^2} [\{c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \cdot \mathbb{1}_{\tau^\Lambda \leq T^1} | y_{1:t}^1] \\ &\quad - \mathbb{E}^{\Lambda, \Gamma^2} [\{c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \cdot \mathbb{1}_{\tau^\Lambda \leq T^1} | y_{1:t}^1] \\ &\quad + \mathbb{E}^{\Lambda_{T^1}, \Gamma^2} [\{c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \cdot \mathbb{1}_{\tau^\Lambda > T^1} | y_{1:t}^1] \\ &\quad - \mathbb{E}^{\Lambda, \Gamma^2} [\{c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \cdot \mathbb{1}_{\tau^\Lambda > T^1} | y_{1:t}^1] \quad (124) \end{aligned}$$

For all realizations where  $\tau^\Lambda \leq T^1$ , the policy  $\Lambda_{T^1}$  has the same stopping time and the final decision as policy  $\Lambda$  and hence they both will send the same realization of messages to O2 and hence O2's policy  $\Gamma^2$  will produce the same realizations of  $\tau^2$  and  $U_{\tau^2}^2$ . This implies that the first two terms in (124) are equal. Thus, (124) becomes

$$\begin{aligned} & \mathbb{E}^{\Lambda_{T^1}, \Gamma^2} [\{c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \cdot \mathbb{1}_{\tau^\Lambda > T^1} | y_{1:t}^1] \\ &\quad - \mathbb{E}^{\Lambda, \Gamma^2} [\{c^2 \tau^2 + J(U_{\tau^2}^2, H)\} \cdot \mathbb{1}_{\tau^\Lambda > T^1} | y_{1:t}^1] \\ &\leq (c^2 \cdot T^2 + L) \cdot \mathbb{E}[\mathbb{1}_{\tau^\Lambda > T^1} | y_{1:t}^1] \\ &= (c^2 \cdot T^2 + L) \cdot P(\tau^\Lambda > T^1 | y_{1:t}^1) \quad (125) \end{aligned}$$

where we used the fact that  $\tau^2$  is bounded by  $T^2$  under policy  $\Gamma^2$  by assumption. Since the stopping time under policy  $\Lambda$  is almost surely finite (otherwise cost of policy would be infinite), we have that  $P(\tau^\Lambda > T^1 | y_{1:t}^1) \rightarrow 0$ , as  $T^1 \rightarrow \infty$ . Thus, from equations (122)-(125), we conclude that for any  $\epsilon > 0$ , there exists a horizon  $T^1$  large enough such that  $W_t(\Lambda_{T^1}) - W_t(\Lambda) \leq \epsilon$ . Therefore,

$$\lim_{T^1 \rightarrow \infty} W_t(\Lambda_{T^1}) = W_t(\Lambda)$$

Hence, we conclude that there does not exist any policy  $\Lambda \in \mathcal{B}^\infty$  for which  $W_t(\Lambda) < \lim_{T^1 \rightarrow \infty} \tilde{V}_t^{T^1}(\bar{\pi}_t^1)$ . Therefore,  $V_t^\infty(y_{1:t}^1) = \lim_{T^1 \rightarrow \infty} V_t^{T^1}(\bar{\pi}_t^1)$ .

#### ACKNOWLEDGMENTS

This research was supported in part by NSF Grant CCR-0325571 and NASA Grant NNX06AD47G.

#### REFERENCES

- [1] R. R. Tenney and N. R. Sandell Jr., "Detection with distributed sensors," *IEEE Trans. Aerospace Electron. Systems*, vol. AES-17, no. 4, pp. 501–510, July 1981.
- [2] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Statistical Signal Processing*. JAI Press, 1993, pp. 297–344.

- [3] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer, 1997.
- [4] R. Radner, "Team decision problems," *The Annals of Math. Statistics*, vol. 33, no. 3, pp. 857–881, Sept. 1962.
- [5] J. N. Tsitsiklis, "Decentralized detection by a large number of sensors," *Mathematics of Control, Signals and Systems*, vol. 1, no. 2, pp. 167–182, 1988.
- [6] J.-F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power-constrained wireless sensor networks," *IEEE Journal on Selected Areas in Communication*, vol. 22, no. 6, pp. 1007–1015, Aug. 2004.
- [7] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Trans. on Info. Theory*, vol. IT-32, no. 4, pp. 533–543, 1986.
- [8] V. V. Veeravalli, T. Basar, and H. Poor, "Decentralized sequential detection with a fusion center performing the sequential test," *IEEE Trans. Inform. Theory*, vol. 39, pp. 433–442, Mar. 1993.
- [9] Y. Mei, "Asymptotic optimality theory for decentralized sequential hypothesis testing in sensor networks," *IEEE Tran. on Info. Theory*, vol. 54, no. 5, pp. 2072–2089.
- [10] G. Fellouris and G. V. Moustakides, "Decentralized sequential hypothesis testing using asynchronous communication," *IEEE Trans. Info. Theory*, submitted.
- [11] D. Teneketzis and Y. C. Ho, "The decentralized wald problem," *Information and Computation*, 73, pp. 23–44, 1987.
- [12] A. LaVigna, A.M. Makowski, and J.S. Baras, "A continuous-time distributed version of the wald's sequential hypothesis testing problem," *Lecture Notes in Control and Information Sciences*, vol. 83, pp. 533–543, 1986.
- [13] A. Wald, *Sequential Analysis*. Wiley, New York, 1947.
- [14] Y. C. Ho, "Team decision theory and information structures," in *Proceedings of the IEEE*, vol. 68, no. 6, 1980, pp. 644–654.
- [15] J. N. Tsitsiklis, "On threshold rules in decentralized detection," in *Proceedings of 25th IEEE Conference of Decision and Control*, Dec. 1986, pp. 232–236.
- [16] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proceedings of the IEEE*, vol. 59, no. 11, pp. 1557–1566, Nov. 1971.
- [17] A. Nayyar and D. Teneketzis, "On the structure of real-time encoders and decoders in a multi-terminal communication system," *IEEE Trans. Info. Theory*, submitted.
- [18] J. Walrand and P. Varaiya, "Optimal causal coding-decoding problems," *IEEE Trans. Inf. theory*, vol. IT-29, no. 6, pp. 814–820, Nov. 1983.