The value of feedback for decentralized detection in large sensor networks

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ISWPC.2011.5751320">http://dx.doi.org/10.1109/ISWPC.2011.5751320</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers (IEEE)</td>
</tr>
<tr>
<td>Version</td>
<td>Author's final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Fri Feb 08 02:46:42 EST 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/73559">http://hdl.handle.net/1721.1/73559</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike 3.0</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a></td>
</tr>
</tbody>
</table>
The Value of Feedback for Decentralized Detection in Large Sensor Networks

Wee Peng Tay, Member, IEEE, John N. Tsitsiklis, Fellow, IEEE

Abstract—We consider the decentralized binary hypothesis testing problem in networks with feedback, where some or all of the sensors have access to compressed summaries of other sensors’ observations. We study certain two-message feedback architectures, in which every sensor sends two messages to a fusion center, with the second message based on full or partial knowledge of the first messages of the other sensors. Under either a Neyman-Pearson or a Bayesian formulation, we show that the asymptotically optimal (in the limit of a large number of sensors) detection performance (as quantified by error exponents) does not benefit from the feedback messages.

Index Terms—Decentralized detection, feedback, error exponent, sensor networks.

I. INTRODUCTION

In the problem of decentralized detection, each sensor makes an observation, and sends a summary of that observation by first applying a quantization function to its observation, and then communicating the result to a fusion center. The fusion center makes a final decision based on all the sensor messages. The goal is to design the sensor quantization functions and the fusion rule so as to minimize a cost function, such as the probability of an incorrect final decision. This problem has been widely studied for various tree architectures, including the parallel configuration [1]–[10], tandem networks [11]–[14], and bounded height tree architectures [15]–[22]. For sensor observations not conditionally independent given the hypothesis, the problem of designing the quantization functions is known to be NP-hard [23]. For this reason, most of the literature assumes that the sensor observations are conditionally independent.

Non-tree networks are harder to analyze because the different messages received by a sensor are not in general conditionally independent. While some structural properties of optimal decision rules are available (see, e.g., [24]), not much is known about the optimal performance. Networks with feedback face the same difficulty. In this paper we consider sensor network architectures that involve feedback: some or all of the sensors have access to compressed summaries of other sensors’ observations. We are interested in characterizing the performance under different architectures, and, in particular, to determine whether the presence of feedback can substantially enhance performance. In the context of a wireless network, feedback can result in unnecessary communication and computation costs, therefore it is important to quantify the performance gain, if any, that feedback can provide.

A variety of feedback architectures, under a Bayesian formulation, have been studied in [25], [26]. These references show that it is person-by-person optimal for every sensor to use a likelihood ratio quantizer, with thresholds that depend on the feedback messages. However, because of the difficulty of optimizing these thresholds when the number of sensors becomes large, it is difficult to analytically compare the performance of networks with and without feedback. Numerical examples in [26] show that a system with feedback has lower probability of error, as expected. To better understand the asymptotics of the error probability, [27] studies the error probability decay rate under a Neyman-Pearson formulation for two different feedback architectures. For either case, it shows that if the fusion center also has access to the feedback messages, then feedback does not improve the optimal error exponent. References [28], [29] consider the Neyman-Pearson problem in a “daisy-chain” architecture, and obtain a similar result.

In this paper, we revisit the two-message architecture studied in [27], and extend the available results. We also study certain feedback architectures that have not been studied before. Our main contributions are as follows.

1) We consider the two-message full feedback architecture studied in [27]. Here, each sensor gets to transmit two messages, and the second message can take into account the first messages of all sensors. We resolve an open problem for the Bayesian formulation, by showing that there is no performance gain over the non-feedback case. We also provide a variant of the result of [27] for the Neyman-Pearson case. Our model is different because unlike [27], because we do not restrict the feedback message alphabet to grow at most subexponentially with the number of sensors.

2) We also study a new two-message sequential feedback architecture. Sensors are indexed, and the second message of a sensor can take into account the first message of all sensors with lower indices. We show that under either the Neyman-Pearson or Bayesian formulation, feedback does not improve the error exponent.

The remainder of this paper is organized as follows. In Section II we define the model, formulate the problems that we will be studying, and provide some background material. In Section III, we study two-message feedback architectures (sequential and full feedback). We offer concluding remarks and discuss open problems in Section IV.
II. PROBLEM FORMULATION

We consider a decentralized binary detection problem involving \( n \) sensors and a fusion center. Each sensor \( k \) observes a random variable \( X_k \) taking values in some measurable space \((\mathcal{X}, \mathcal{F})\), and is distributed according to a measure \( \mathbb{P}_j \) under hypothesis \( H_j \), for \( j = 0, 1 \). Under either hypothesis \( H_j \), \( j = 0, 1 \), the random variables \( X_k \) are assumed to be i.i.d. We use \( \mathbb{E}_f \) to denote the expectation operator with respect to \( \mathbb{P}_j \).

A. Feedback Architectures

In the two-message feedback architecture (see Figure 1), each sensor \( k \) sends a message \( Y_k = \gamma_k(X_k) \), which is a quantized version of its observation \( X_k \), to the fusion center. The quantization function is of the form \( \gamma_k: \mathcal{X} \rightarrow \mathcal{T} \), where \( \mathcal{T} \) is the transmission signal set. In most engineering scenarios, \( \mathcal{T} \) is assumed to be a finite alphabet, although we do not require this restriction. We denote by \( \Gamma \) the set of allowable quantization functions.

![Fig. 1. A two-message architecture.](image)

We assume that the sensors are indexed in order of when they send their messages to the fusion center. We will consider two forms of feedback under the two-message architecture. In the first form, the information \( W_k = (Y_1, \ldots, Y_{k-1}) \) received from sensors \( 1, \ldots, k-1 \), is fed back by the fusion center to sensor \( k \). We call this sequential feedback. In the second form of feedback, the fusion center broadcasts the messages \( (Y_1, \ldots, Y_n) \) to all the sensors. In this case, the additional information received at sensor \( k \) is \( W_k = (Y_1, \ldots, Y_{k-1}, Y_{k+1}, \ldots, Y_n) \). We call this full feedback.

In both feedback scenarios, each sensor forms a new second message \( Z_k = \delta_k(X_k, W_k) \) based on the additional information \( W_k \), and sends it to the fusion center. Finally, the fusion center makes a decision \( Y_f = \gamma_f(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n) \). The collection \( (\gamma_f, \gamma_1, \ldots, \gamma_n, \delta_1, \ldots, \delta_n) \) is called a strategy. A sequence of strategies, one for each value of \( n \), is called a strategy sequence. Typically, we wish to design strategy sequences that optimize a given objective function.

B. Preliminaries

In this section, we list the basic assumptions that we will be making throughout this paper, and some preliminary mathematical results that we will apply in our subsequent proofs.

Consider the Radon-Nikodym derivative \( \ell_{ij}^X \) of the measure \( \mathbb{P}_i \) with respect to (w.r.t.) the measure \( \mathbb{P}_j \). Informally, this is the likelihood ratio associated with an observation of \( X \). It is a random variable whose value is determined by \( X \); accordingly, its value should be denoted by a notation such as \( \ell_{ij}^X(X) \).

We consider the Radon-Nikodym derivative of the conditional distribution of \( X \) given \( Y \) under \( \mathbb{P}_i \) w.r.t. that under \( \mathbb{P}_j \).

For technical reasons (see [30]), we make the following assumptions.

Assumption 1: The measures \( \mathbb{P}_0 \) and \( \mathbb{P}_1 \) are absolutely continuous w.r.t. each other. Furthermore, there exists some \( \gamma \in \Gamma \) such that \( -\mathbb{E}_0[\log \ell_{01}(\gamma(X_1))] < 0 < \mathbb{E}_1[\log \ell_{10}(\gamma(X_1))] \).

Assumption 2: We have \( \mathbb{E}_0[\log^2 \ell_{01}(X_1)] < \infty \) and \( \mathbb{E}_1[\log^2 \ell_{10}(X_1)] < \infty \).

We will require the following lower bound for the maximum of the Type I and II error probabilities. This bound was first proved in [31] for the case of discrete observation space. The following proposition generalizes the result to a general observation space. The proof is almost identical to that in [31], and is omitted here.\(^1\)

**Proposition 1:** Consider a hypothesis testing problem based on a single observation \( X \) with distribution \( \mathbb{P}_j \) under hypothesis \( H_j, j = 0, 1 \). Let \( P_{e,0}^{(ij)} \) be the probability of error when \( H_j \) is true. Let \( Z = \log \frac{d\mathbb{P}_0}{d\mathbb{P}_j}(X) \) be the log-Radon-Nikodym derivative. For any \( s \in \mathbb{R} \), let \( \Lambda(s) = \log \mathbb{E}_0[\exp(sZ)] \) be the log-moment generating function of \( Z \). Then, for \( s^* \in [0, 1] \) such that \( \Lambda'(s^*) = 0 \), we have

\[
\max(P_{e,0}, P_{e,1}) \geq \frac{1}{4} \exp \left( \Lambda(s^*) - \sqrt{2\Lambda''(s^*)} \right).
\]

III. ASYMPTOTIC PERFORMANCE

In this section, we study the Neyman-Pearson and Bayesian formulations of the decentralized detection problem in the two-message architecture. For both sequential and full feedback, and \( i, j \in \{0, 1\} \), let the log likelihood ratio at the fusion center be

\[
\mathcal{L}_{ij}^{(n)} = \log \ell_{ij}(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n) = \sum_{k=1}^{n} \log \ell_{ij}(Y_k) + \log \ell_{ij}(Z_1, \ldots, Z_n|Y_1, \ldots, Y_n).
\]

To simplify notation, we let

\[
\mathcal{L}_{ij}^{k}(w) = \log \ell_{ij}(Y_k) + \log \ell_{ij}(Z_k|Y_k, w) = \log \ell_{ij}(\gamma_k(X_k), \delta_k^w(X_k)),
\]

where \( \delta_k^w : \mathcal{X} \rightarrow \mathcal{T} \) is a function that depends on the value \( w \).

\(^1\)Throughout this paper, we use \( f'(s) \) to denote the derivative of \( f \) w.r.t. \( s \).
A. Neyman-Pearson Formulation

Let $\alpha \in (0, 1)$ be a given constant. A strategy is admissible if its Type I error probability satisfies $P_0(Y_f = 1) < \alpha$. Let $\beta_{n}^{*} = \inf P_1(Y_f = 0)$, where the infimum is taken over all admissible strategies. Our objective is to characterize the optimal error exponent $\limsup_{n \to \infty} (1/n) \log \beta_{n}^{*}$ under different feedback architectures.

Let $g_{p}^{*}$ be the optimal error exponent of the parallel configuration in which there is no feedback from the fusion center, i.e., each sensor $k$ sends a message $(\gamma_k(X_k), \delta_k(X_k))$ to the fusion center. From [30], the optimal error exponent is

$$g_{p}^{*} = \sup_{(\gamma, \delta) \in \Gamma^2} \mathbb{E}_0[\log \ell_0(\gamma(X_1), \delta(X_1))]$$

Let $g_{sf}^{*}$ and $g_{f}^{*}$ be the optimal error exponents for the sequential feedback and full feedback architectures respectively. Since the sensors can ignore some or all of the feedback messages from the fusion center, we have

$$g_{f}^{*} \leq g_{sf}^{*} \leq g_{p}^{*}. \quad (1)$$

We will show that under appropriate but mild assumptions, the inequalities in (1) are equalities. Hence, from an asymptotic view point, both sequential and full feedback results in no gain in detection performance. This is in line with the results in [29], which shows that feedback does not improve the optimal error exponent in a diary chain architecture. We first show an instructive result that will guide us in the main proofs.

**Lemma 1:** Suppose that $\beta_{n}$ is the Type II error probability of a given strategy. If $R > 0$ and

$$\limsup_{n \to \infty} P_0(L_{01}^{(n)} > nR) < 1 - \alpha,$$

then

$$\liminf_{n \to \infty} \frac{1}{n} \log \beta_{n} \geq -R.$$

**Proof:** We have

$$\beta_{n} = \mathbb{P}_1(Y_f = 0) = \mathbb{E}_0\left[\exp(-L_{01}^{(n)})\mathbb{1}_{[Y_f = 0]}\right] \geq \mathbb{E}_0\left[\exp(-L_{01}^{(n)})\mathbb{1}_{[Y_f = 0, L_{01}^{(n)} \leq nR]}\right] \geq e^{-nR}P_0(Y_f = 0, L_{01}^{(n)} \leq nR).$$

Therefore,

$$P_0(Y_f = 0, L_{01}^{(n)} \leq nR) \leq \beta_{n}e^{nR}.$$  

This upper bound yields

$$1 - \alpha \leq P_0(Y_f = 0) \leq \beta_{n}e^{nR} + P_0(L_{01}^{(n)} > nR)$$

and

$$\frac{1}{n} \log \beta_{n} + R \geq \frac{1}{n} \log(1 - \alpha - P_0(L_{01}^{(n)} > nR)).$$

The lemma then follows by letting $n \to \infty$.

**Theorem 1:** Suppose Assumptions 1-2 hold. Then, the optimal error exponent for the sequential feedback architecture is $g_{sf}^{*} = g_{p}^{*}$. Moreover, there is no loss in optimality if sensors ignore the feedback messages from the fusion center, and all sensors are constrained to using the same quantization function.

**Proof:** From (1), we have $g_{sf}^{*} \leq g_{p}^{*}$. To show the reverse inequality, we will first upper bound $E_0[L_{01}^{k}(W_k) \mid W_k]$ by $R = -g_{p}^{*}$, and then apply Lemma 1 to obtain a lower bound for $g_{sf}^{*}$.

**Proof:** From (1), we have $g_{sf}^{*} \leq g_{p}^{*}$. To show the reverse inequality, we will first upper bound $E_0[L_{01}^{k}(W_k) \mid W_k]$ by $R = -g_{p}^{*}$, and then apply Lemma 1 to obtain a lower bound for $g_{sf}^{*}$. We have for all $w$,

$$E_0[L_{01}^{k}(W_k) \mid W_k = w] = \mathbb{E}_0[\log \ell_0(\gamma_k(X_k), \delta_k^{w}(X_k)) \mid W_k = w] \leq \sup_{(\gamma, \delta) \in \Gamma^2} \mathbb{E}_0[\log \ell_0(\gamma(X_1), \delta(X_1))] = R. \quad (2)$$

From Assumption 2, there exists some constant $a > 0$ such that

$$\text{var}_0(L_{01}^{k}(W_k)) \leq \mathbb{E}_0\left[(L_{01}^{k}(W_k))^2 \mid W_k\right] \leq a. \quad (3)$$

Recall that $W_k = (Y_1, \ldots, Y_{k-1})$, so we have, for $m < k$,

$$E_0[(L_{01}^{m}(W_m) - E_0[L_{01}^{m}(W_m) \mid W_m]) \cdot (L_{01}^{k}(W_k) - E_0[L_{01}^{k}(W_k) \mid W_k])] = E_0[(L_{01}^{m}(W_m) - E_0[L_{01}^{k}(W_k) \mid W_k])] \cdot E_0[L_{01}^{k}(W_k) - E_0[L_{01}^{k}(W_k) \mid W_k)] = 0. \quad (4)$$

Let $\epsilon > 0$. From Chebychev’s inequality, together with (2), (3), and (4), we obtain

$$\mathbb{P}_0 \left(L_{01}^{(n)} > n(1 + \epsilon)R\right) \leq \mathbb{P}_0 \left(\sum_{k=1}^{n} (L_{01}^{(n)} - E_0[L_{01}^{k}(W_k) \mid W_k]) > n\epsilon R\right) \leq \frac{a}{n\epsilon^2 R^2}.$$  

Letting $n \to \infty$, we get

$$\lim_{n \to \infty} \mathbb{P}_0 \left(L_{01}^{(n)} > n(1 + \epsilon)R\right) = 0.$$  

Therefore, applying Lemma 1, we have $g_{sf}^{*} \geq (1 + \epsilon)g_{p}^{*}$. Since $\epsilon$ was chosen arbitrarily, we have $g_{sf}^{*} \geq g_{p}^{*}$, and the proof is now complete.

Next, we consider the full feedback architecture. The same architecture has been studied in [27], using the method of types. In the following, we show the same result as [27], i.e., there is no gain from the feedback messages asymptotically, but without the constraint that the feedback messages have an alphabet that grows at most subexponentially fast.

For technical reasons, we will require the following additional assumption.

**Assumption 3:** Let $b(\lambda) = \log \mathbb{E}_0[\exp(\lambda \log \ell_0(X_1))]$. There exists $s > 0$ such that $b(s) < \infty$.

Note that since $b(\cdot)$ is non-decreasing on $[0, s]$ (see Lemma 2.2.5 of [32]), and therefore Assumption 3 implies that $b(\lambda) < \infty$ for all $\lambda \in [0, s]$. 
Suppose that a strategy sequence has been fixed. Let \( \varphi_n(\lambda) = \log \mathbb{E}_0[\exp(\lambda L_0^{(n)} / n)] \). The Fenchel-Legendre transform of \( \varphi_n \) is \( \Phi_n(t) = \sup_{\lambda \in \mathbb{R}} \{At - \varphi_n(\lambda)\} \). We first show some properties of \( \varphi_n \), and use these properties together with Lemma 1 to prove that similar to the sequential feedback architecture, feedback does not improve the asymptotic performance of a full feedback architecture. Denote by \( \varphi_n'(\lambda) \) the derivative of \( \varphi_n \) w.r.t. \( \lambda \).

**Lemma 2:** Suppose Assumption 3 holds, and let \( s \) be as in the assumption. Suppose \( \lambda \in [0, s) \). Then, for \( n > 1 \), we have \( 0 \leq \varphi_n(\lambda) \leq nb(\lambda/n), 0 \leq \varphi_n'(\lambda) \leq C_1, \) and \( 0 \leq \varphi_n''(\lambda) \leq C_2 \) for some constants \( C_1, C_2 > 0 \), independent of the strategy used and \( n \). Furthermore, \( \varphi_n(0) \leq -g_p^* \) for all \( n \).

**Proof:** Since \( \varphi_n \) is a convex function [32] with \( \varphi_n(-1) = \varphi_n(0) = 0, \varphi_n(\lambda) \geq 0 \) for \( \lambda \geq 0 \). Using Jensen’s inequality, we obtain

\[
\frac{1}{n} \varphi_n(\lambda) = \frac{1}{n} \log \mathbb{E}_0 \left[ \exp(\lambda L_0^{(n)} / n) \right] \\
\leq \frac{1}{n} \log \mathbb{E}_0 \left[ \exp(\lambda/n \log \ell_0(X_1, \ldots, X_n)) \right] \\
= \frac{1}{n} \sum_{k=1}^n \log \mathbb{E}_0 \left[ \exp((\lambda/n) \log \ell_0(X_k)) \right] \\
= b(\lambda/n) < \infty.
\]

(5)

From (5), it can be shown that \( \varphi_n \) is twice differentiable throughout \( [0, s) \) (see e.g. Example A.5.2 of [33]). To simplify the notation in the following proof, let \( \ell_n = \ell_0(Y_1, \ldots, Y_n, Z_1, \ldots, Z_n) \) and \( \ell_{1,n} = \ell_0(X_1, \ldots, X_n) \). Since \( \varphi_n(\lambda) \geq 0 \), we have \( \varphi_n'(\lambda) \geq 0 \).

**Proof:**

\[
\varphi_n'(\lambda) = \frac{1}{n} e^{-\varphi_n(\lambda)} \mathbb{E}_0 \left[ e^{(\lambda/n) L_0^{(n)} / n} L_0^{(n)} \right] \\
\leq \frac{1}{n} \mathbb{E}_0 \left[ (\ell_{1,n})^{\lambda/n} \log \ell_{1,n} \right] \\
= \frac{1}{n} \mathbb{E}_1 \left[ (\ell_{1,n})^{\lambda/n+1} \log \ell_{1,n} \right] \\
\leq \frac{1}{n} \mathbb{E}_1 \left[ \phi(\ell_{1,n}) \right],
\]

(6)

where \( \phi(x) = (x^{\lambda/n+1} \log x) 1_{x \geq 2} \) is a convex function. Since \( \ell_{1,n} = \mathbb{E}_1[\ell_{1,n} | F] \), where \( F \) is the \( \sigma \)-field generated by \( Y_k \) and \( Z_k \), \( k = 1, \ldots, n \), we get from Jensen’s inequality that the R.H.S. of (6) is bounded above by

\[
\frac{1}{n} \mathbb{E}_1 \left[ \phi(\ell_{1,n}) \right] \\
= \frac{1}{n} \mathbb{E}_0 \left[ (\ell_{1,n})^{\lambda/n} \log \ell_{1,n} \right] \\
= \frac{1}{n} \mathbb{E}_0 \left[ (\ell_{0}(X_1))^{\lambda/n} \log \ell_{0}(X_k) \right] \\
= \mathbb{E}_0 \left[ (\ell_{0}(X_1))^{2\lambda/n} \log \ell_{0}(X_1) \right] \mathbb{E}_0 \left[ (\ell_{0}(X_1))^{\lambda/n} \right] \mathbb{E}_0 \left[ (\ell_{0}(X_1))^{1-\lambda/n} \right] \\
\leq \mathbb{E}_0 \left[ (\ell_{0}(X_1))^{2\lambda/n} \log \ell_{0}(X_1) \right] \mathbb{E}_0 \left[ (\ell_{0}(X_1))^{\lambda/n} \right] \mathbb{E}_0 \left[ (\ell_{0}(X_1))^{1-\lambda/n} \right]
\]

where the last equality is because the \( X_i \) are i.i.d., and the last inequality follows from the Cauchy-Schwarz inequality and Jensen’s inequality. From Assumption 3, the R.H.S. of (7) is bounded by some constant \( C_1 \), independent of \( n \) and \( \lambda \). The proof for the bound on the second derivative is similar, and is omitted. Finally, the bound for \( \varphi_n'(0) = (1/n) \mathbb{E}_0 \left[ L_0^{(n)} \right] \) follows from the same proof as for (2). The proof of the lemma is now complete.

**Lemma 3:** Suppose Assumptions 1-3 hold, and \( t > r = \limsup_{n \to \infty} \varphi_n'(0) \). Then, there exist \( N_0 \) and \( \eta > 0 \) such that for all \( n > N_0, \varphi_n(t) \geq \eta \).

**Proof:** Let \( \epsilon = t - r > 0 \), and set \( \lambda = \min(s/2, \epsilon/(2C_2)) > 0 \). Then, from Lemma 2, we have \( \varphi_n(\lambda) \leq (\varphi_n(0) + \lambda C_2) \lambda \). There exists \( N_0 \) such that for all \( n > N_0, \varphi_n'(0) \leq r + \epsilon/4 \). Therefore, we get

\[
\varphi_n(\lambda) \leq (r + 3\epsilon/4) \lambda \\
= (t - \epsilon/4) \lambda.
\]

Therefore, \( \Phi_n(t) \geq \lambda t - \varphi_n(\lambda) \geq \epsilon \lambda/4 \). The proof of the lemma is now complete.

Finally, we can prove the following result.

**Theorem 2:** Suppose Assumptions 1-3 hold. Then, in the full feedback architecture, there is no loss in optimality if sensors ignore the feedback messages from the fusion center, i.e., \( g^{\ast}_f = g^*_p \). Moreover, there is no loss in optimality if all sensors are constrained to using the same quantization function.

**Proof:** Let \( \epsilon > 0 \) and \( t = -(1 + \epsilon) g^*_p \). From Lemma 2,

\[
\limsup_{n \to \infty} \varphi_n'(0) = -g^*_p.
\]

Applying the Chernoff bound and Lemma 3, we have for \( n \) sufficiently large,

\[
\frac{1}{n} \log \mathbb{P}_0 \left( L_0^{(n)} > nt \right) \leq -\Phi_n(t) \leq -\eta,
\]

where \( \eta > 0 \) is as chosen in Lemma 3. Therefore we can apply Lemma 1 to get \( g^{\ast}_f \geq (1 + \epsilon) g^*_p \). Taking \( \epsilon \to 0 \) and applying (1), we obtain the theorem.

**B. Bayesian Formulation**

Let the prior probability of hypothesis \( H_j \) be \( \pi_j > 0, j = 0, 1 \). Given a strategy, the probability of error at the fusion center is \( P_e = \pi_0 \mathbb{P}_0(Y_j = 1) + \pi_1 \mathbb{P}_1(Y_j = 0) \). Let \( P^*_e \) be the minimum probability of error, over all strategies. We seek to characterize the optimal error exponent

\[
\limsup_{n \to \infty} \frac{1}{n} \log P^*_e.
\]

From [30], the optimal error exponent for the parallel configuration without any feedback is given by

\[
E_p^* = \inf_{(C, \delta) \in \mathbb{T}^2} \min_{\lambda \in [0, 1]} \log \mathbb{E}_0 \left[ \exp(\lambda \log \ell_0(\gamma(X_1), \delta(X_1))) \right].
\]

Similar to the Neyman-Pearson case, we let \( E^*_a \) and \( E^*_f \) denote the optimal error exponent for the sequential feedback and full feedback architectures respectively. We first study the full feedback architecture, and infer results about the sequential feedback architecture.

**Lemma 4:** Suppose Assumptions 1 and 2 hold.
Recall that $\epsilon > 0$.

For all $s \in (0, 1)$ with $\psi_n(s_n) = t$. Then, there exists a constant $C$ such that for all $n$, we have $\psi_n(s_n) \leq nC$.

(iii) For all $s \in [0, 1]$, we have $\psi_n(s) \geq n \psi_s^*$.

Proof: (Outline) Claims (i) and (ii) follow from calculus. We omit their proofs here. In the following, we show claim (iii). Let $\lambda = \frac{\ln n}{\ln \lambda}$. Then $\lambda \in \gamma$ and $Y_n = (Y_1, \ldots, Y_n)$. We have

$$\psi_n(\lambda) = \log E_0 \left[ (\ell_0(Y_n)^\lambda \right] \cdot E_0 \left[ \prod_{k=1}^n (\ell_0(Z_k | Y_k, W_k)^\lambda \right] | Y_n^\lambda \right].$$

(8)

Recall that $Z_k = \delta_k^W(X_k)$, where $W_k = (Y_{k-1}, Y_n^k)$ in the full feedback configuration. Consider the inner expectation on the R.H.S. of (8). For $\epsilon > 0$, we have

$$E_0 \left[ \prod_{k=1}^n (\ell_0(Z_k | Y_k, W_k)^\lambda \right] | Y_n^\lambda \right]$$

$$= E_0 \left[ \prod_{k=1}^n (\ell_0(\delta_k^W(X_k) | Y_k)^\lambda \right] | Y_n^\lambda \right]$$

$$= \prod_{k=1}^n E_0 \left[ (\ell_0(\delta_k^W(X_k) | Y_k)^\lambda \right] | Y_k^\lambda \right]$$

$$\geq \prod_{k=1}^n \left( E_0 \left[ (\ell_0(\delta_k^W(X_k) | Y_k)^\lambda \right] | Y_k - \epsilon \right),$$

where $\delta_k^W(X_k)$ is a function depending on the value of $Y_k$, and is such that

$$E_0 \left[ (\ell_0(\delta_k^W(X_k) | Y_k)^\lambda \right] | Y_k \right]$$

$$\leq \inf_{\delta \in \Gamma} E_0 \left[ (\ell_0(\delta(X_k) | Y_k)^\lambda \right] | Y_k + \epsilon \right].$$

From (8), we obtain

$$\psi_n(\lambda) \geq \log E_0 \left[ (\ell_0(Y_n)^\lambda \right] \cdot \prod_{k=1}^n \left( E_0 \left[ (\ell_0(\delta_k^W(X_k) | Y_k)^\lambda \right] | Y_k - \epsilon \right] \right]$$

$$= \sum_{k=1}^n \log E_0 \left[ (\ell_0(Y_k)^\lambda \right] \cdot \left( E_0 \left[ (\ell_0(\delta_k^W(X_k) | Y_k)^\lambda \right] | Y_k - \epsilon \right] \right]$$

$$\geq \sum_{k=1}^n \log E_0 \left[ (\ell_0(Y_k, \gamma_k^W(X_k))^\lambda \right] | Y_k - \epsilon \right] \right],$$

(9)

where we used the inequality $E_0 \left[ (\ell_0(Y_k)^\lambda \right] \leq 1$ in (9).

Recall that $Y_k = \gamma_k(X_k)$. We can define $\xi_k \in \Gamma^2$ such that $\xi_k(X_k) = (\gamma_k(X_k), \delta_k(X_k))$, where $\delta_k(X_k) = \delta_k^W(X_k)$ if $\gamma_k(X_k) = u \in \mathcal{T}$. From (9), we obtain the bound

$$\psi_n(\lambda) \geq \sum_{k=1}^n \log \left( E_0 \left[ (\ell_0(\xi_k(X_k))^\lambda \right] - \epsilon \right)$$

$$\geq n \log \left( \inf_{\xi \in \Gamma^2} E_0 \left[ (\ell_0(\xi(X_1))^\lambda \right] - \epsilon \right)$$

Since $\epsilon$ is arbitrary, we have

$$\psi_n(\lambda) \geq n \inf_{\xi \in \Gamma^2} \log E_0 \left[ (\ell_0(\xi(X_1))^\lambda \right]$$

$$\geq n \psi_s^*,$$

and the lemma is proved.

Theorem 3: Suppose Assumptions 1 and 2 hold. Then $\mathcal{E}_s^* = \mathcal{E}_p^*$. Moreover, there is no loss in optimality if sensors are constrained to using the same quantization function, which ignore the feedback messages from the fusion center.

Proof: It is clear that $\mathcal{E}_s^* \leq \mathcal{E}_p^*$. To show the reverse bound, we make use of Proposition 1. Let the conditional probability of error under $H_j$ be $P_{nj}$ for $j = 0, 1$. Let $s_n^* = \arg \min_{s \in (0, 1)} \psi_n(s)$ so that $\psi_n(s_n^*) = 0$. From Proposition 1, we have

$$\max_{j=0, 1} P_{nj} \geq \frac{1}{4} \exp \left( \psi_n(s_n^*) - \sqrt{2 \psi_n(s_n^*)} \right)$$

$$\geq \exp \left( \psi_n(s_n^*) \right) - C \sqrt{n}$$

$$\geq \exp \left( n \psi_s^* \right) - C \sqrt{n},$$

where $C$ is some constant. In the above, the penultimate inequality follows from Lemma 4(ii), and the last inequality from Lemma 4(iii). Letting $n \to \infty$, we have

$$\limsup_{n \to \infty} \frac{1}{n} \log P_e = \limsup_{n \to \infty} \frac{1}{n} \log \max_{j=0, 1} P_{nj}$$

$$\geq \mathcal{E}_p^*.$$

This implies that $\mathcal{E}_s^* \geq \mathcal{E}_p^*$, and the theorem is proven.

Since the sequential feedback configuration can perform no better than the full feedback architecture, and no worse than the parallel configuration, we have the following result.

Theorem 4: Suppose Assumptions 1 and 2 hold. Then $\mathcal{E}_s^* = \mathcal{E}_p^*$. Moreover, there is no loss in optimality if sensors are constrained to using the same quantization function, which ignore the feedback messages from the fusion center.

IV. Conclusion

We have studied the two-message feedback architecture, in which each sensor has access to compressed summaries of some or all other sensors’ first messages to the fusion center. In the sequential feedback architecture, each sensor has access to the first messages of those sensors that communicate with the fusion center before it. In the full feedback architecture, each sensor has access to the first messages of every other sensor. In both architectures, and under both the Neyman-Pearson and Bayesian formulations, we show that the optimal error exponent is not improved by the feedback messages. Our results suggest that in the regime of a large number of sensors, the performance gain due to feedback does not justify the increase in communication and computation costs incurred in a feedback architecture.

Future research directions include studying the effect of feedback in architectures where all sensors send only one message to the fusion center, and where a group of sensors has access to the messages of the other sensors. Other more complex feedback architectures like hierarchial networks with
feedback messages from one level to the next are also important future research directions. We would also like to consider the impact of feedback on distributed multiple hypothesis testing and parameter estimation.

REFERENCES


