

AN INFORMATION PROCESSING APPROACH TO DISTRIBUTED DETECTION

Michael A. Lexa and Don H. Johnson

Rice University
Department of Electrical and Computer Engineering
Houston, TX 77005-1892
amlexa@rice.edu, dhj@rice.edu

ABSTRACT

We apply the recent theory of information processing to a hybrid distributed detection architecture that combines the traditional parallel and tandem architectures. Central to this theory is the Kullback-Leibler discrimination distance and a quantity known as the information transfer ratio, defined as the ratio of the KL distances between the distributions characterizing the input and output of a system. We characterize the asymptotic performance of a proposed hybrid system and compare it with the performance of the parallel, tandem and centralized architectures. We conclude with an illustrative example.

1. INTRODUCTION

Distributed detection is a basic signal processing task in the emerging field of sensor networks, and despite the tremendous amount of existing distributed detection literature, there is a sizeable gap between known results and the proposed capabilities of the next generation networks [1]. While this paper does not specifically address any of these shortcomings, we introduce a new approach to distributed detection which we hope will prove useful in solving these new challenges.

We apply the theory of information processing [2] to a practical distributed detection architecture which combines the standard parallel and tandem architectures. (See [3, 4] and the references therein for a complete review of the parallel and tandem architectures.) Central to this theory is the Kullback-Leibler (KL) discrimination distance [5] and a quantity known as the *information transfer ratio*. The information transfer ratio, denoted by γ , is defined as the ratio of the KL distances between the distributions characterizing the input and output of a system. In a very broad sense, the theory of information processing endeavors to quantify how systems process information. Here, we narrow our attention to the implications of Stein's Lemma [5, 6] which states that the KL distance is the asymptotic error rate of an optimal Neyman-Person detector. It is through this lemma that γ primarily gains its meaning.

The standard parallel system consists of a set of detectors and a central processing unit or fusion center. Each detector collects observations and based upon these data, decides upon the active hypothesis. The decisions are then transmitted to the fusion center which, in turn, makes a final global decision. The tandem architecture, on the other hand, has no fusion center. The detectors are ordered, and they relay their decisions to each other. The process begins when the first detector makes a decision based on its observation. This first decision is then transmitted to the second detector. Based on the first detector's decision and its own observation, the second detector makes a decision which is passed on to the third detector. The process repeats, with the decision of the last detector taken to be the final decision of the system. In the combined or hybrid architecture decisions are relayed among neighboring detectors as well as transmitted to the fusion center. See figure 1.

Throughout this paper, we assume the fusion center performs a Neyman-Pearson likelihood ratio test. That is, we assume one of the fusion center's error probabilities is constrained and the other maximized. However, instead of approaching this optimization directly using Lagrange multipliers, we parameterize the decision rules and examine the KL distance between the joint distributions under each hypothesis at the *input* of the fusion center. Because of Stein's Lemma, we know the KL distance, computed at the input of the fusion center, is directly linked to the fusion center's (and therefore the entire hybrid system's) performance. The greater the KL distance, the better the system will be able to discriminate one hypothesis from the other.

To simplify our calculations, we choose to use a common decision rule for all of the local detectors. In particular, we choose to use a variant of what Tang *et al.* termed the two-sided constant control strategy (CCS) [7]. Letting Y_n denote the input observation of the n^{th} detector under each

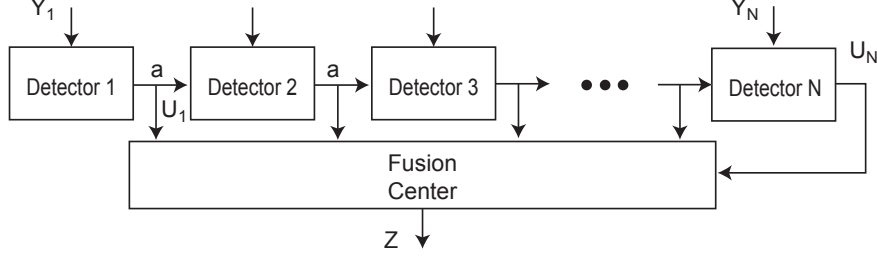


Figure 1: This is the hybrid distributed detection architecture. All the detectors, except the first, make their decisions based upon their observations and a scaled version of the preceding detector's decision. The parameter a controls the scaling of the decisions. The decision of the fusion center is the final, global decision. If the fusion center simply blocks the first $(N - 1)$ inputs but passes the N^{th} input to its output, the hybrid system reduces to the standard tandem architecture. When $a = 1$, there is effectively no interactions between the detectors and the hybrid system becomes the parallel architecture.

hypothesis H_1 and H_0 , the decision rule is given by

$$\Lambda(Y_1) \underset{H_0}{\overset{H_1}{\geq}} t \quad (1)$$

$$\Lambda(Y_n) \underset{H_0}{\overset{H_1}{\geq}} ta^{U_{n-1}}, \quad 2 \leq n \leq N \quad (2)$$

where $\Lambda(\cdot)$ is the likelihood ratio, t is a baseline threshold, $a > 0$ is a linking parameter, and $U_{n-1} \in \{-1, 1\}$ is the binary decision of the $(n - 1)^{\text{th}}$ detector. In words, the detectors operate and interact in the following way. The n^{th} ($n \geq 2$) detector collects an observation y_n and evaluates the likelihood ratio at that point. If $U_{n-1} = 1$, $\Lambda(y_n)$ is compared to the baseline threshold scaled by a ; if $U_{n-1} = -1$, it is compared to $\frac{t}{a}$. In either case, if $\Lambda(Y_n)$ is greater than its threshold then $U_n = 1$, otherwise $U_n = -1$. Tang *et al.* proposed the two-sided CCS as a suboptimal decision strategy for the tandem architecture. They showed, however, that this strategy performs nearly as well as the optimal centralized detector in terms of probability of error. Because of this fact and because of its simplicity, we use it here in the hybrid architecture.

Using the joint probability density functions $p(Y_1, Y_2, \dots, Y_N | H_i) = p_i(\mathbf{Y})$ and $p(U_1, U_2, \dots, U_N | H_i) = p_i(\mathbf{U})$ we can express the information transfer ratio across the tandem portion of the hybrid architecture in one of two ways.

$$\gamma_{01} = \frac{\mathcal{D}(p_0(\mathbf{U}) \| p_1(\mathbf{U}))}{\mathcal{D}(p_0(\mathbf{Y}) \| p_1(\mathbf{Y}))} \quad \gamma_{10} = \frac{\mathcal{D}(p_1(\mathbf{U}) \| p_0(\mathbf{U}))}{\mathcal{D}(p_1(\mathbf{Y}) \| p_0(\mathbf{Y}))}$$

where for two continuous density functions $p_0(x), p_1(x)$ the KL distance between $p_0(x)$ relative to $p_1(x)$ is defined as the expected value of the likelihood ratio with respect to p_0 .

$$\mathcal{D}(p_0 \| p_1) = \int p_0(x) \log \frac{p_0(x)}{p_1(x)} dx$$

We distinguish between the two types of KL distances in the expressions above because, in general, KL distances are not

symmetric ($\mathcal{D}(p_1 \| p_0) \neq \mathcal{D}(p_0 \| p_1)$). The distance to p_0 relative to p_1 is not, in general, the same as the distance to p_1 relative to p_0 . For the remainder of the paper, however, we suppress the subscripts and focus on γ_{01} because it's the pertinent quantity when we fix the fusion center's false alarm probability and try to maximize its probability of detection. Due to the Data Processing Theorem (DPT) [6], we know $\mathcal{D}(p_0(\mathbf{Y}) \| p_1(\mathbf{Y})) \geq \mathcal{D}(p_0(\mathbf{U}) \| p_1(\mathbf{U}))$, causing γ to be a number between zero and one. Thus, we interpret γ as being the fractional loss of discrimination distance across the tandem portion of the hybrid system. In light of Stein's Lemma, a γ of $\frac{1}{2}$, for instance, effectively means that the number of independent and identically distributed (iid) data needed to achieve the same performance as the optimal centralized detector doubles and affects the error probability as the square root.

2. INFORMATION TRANSFER RATIO CALCULATIONS

Because we assume the inputs are iid under each hypothesis the KL distance of the joint distributions of \mathbf{Y} is simply the sum of the KL distances of Y_n .

$$\begin{aligned} \mathcal{D}(p_0(\mathbf{Y}) \| p_1(\mathbf{Y})) &= \sum_{n=1}^N \mathcal{D}(p_0(Y_n) \| p_1(Y_n)) \\ &= N \cdot \mathcal{D}(p_0(Y_1) \| p_1(Y_1)) \end{aligned} \quad (3)$$

Using the Markov formulation of the KL distance [2], the distance between the joint distributions of the binary decisions is

$$\begin{aligned} \mathcal{D}(p_0(\mathbf{U}) \| p_1(\mathbf{U})) &= \\ \mathcal{D}(p_0(U_1) \| p_1(U_1)) &+ \sum_{n=2}^N \mathcal{D}(p_0(U_n | U_{n-1}) \| p_1(U_n | U_{n-1})) \end{aligned} \quad (4)$$

As a consequence of the constant control decision strategy, all of the conditional KL distances $\mathcal{D}(p_1(U_n|U_{n-1})||p_0(U_n|U_{n-1}))$ are equal. Thus, (4) reduces to

$$\mathcal{D}(p_0(U_1)||p_1(U_1)) + (N-1)\mathcal{D}(p_0(U_n|U_{n-1})||p_1(U_n|U_{n-1})) \quad (5)$$

The conditional probabilities $p(U_n|U_{n-1}|H_i), i = 0, 1$ depend on the input distributions and are a function of a .

Taking the ratio of (3) and (5), we have

$$\gamma(N, a) = \frac{\mathcal{D}(p_0(U_1)||p_1(U_1))}{N \cdot \mathcal{D}(p_0(Y_1)||p_1(Y_1))} + \frac{(N-1)\mathcal{D}(p_0(U_n|U_{n-1})||p_1(U_n|U_{n-1}))}{N \cdot \mathcal{D}(p_0(Y_1)||p_1(Y_1))} \quad (6)$$

3. ANALYSIS AND RESULTS

We begin with a few general observations. First, letting $N \rightarrow \infty$ in (6), we see that asymptotically, γ tends towards the ratio of the conditional KL distance and the input KL distance.

$$\gamma(a) \stackrel{N \rightarrow \infty}{=} \frac{\mathcal{D}(p_0(U_n|U_{n-1})||p_1(U_n|U_{n-1}))}{\mathcal{D}(p_0(Y_1)||p_1(Y_1))} \quad (7)$$

This simple result allows us to identify diminishing points of return as N increases, i.e. identify the point where adding detectors no longer results in system gain.

Second, when $a = 1$, there are no interactions between the detectors and the hybrid architecture becomes the traditional parallel architecture. The output marginal distributions $p_i(U_n), i = 0, 1$ become iid and the KL distance between the joint distributions $p_i(\mathbf{U})$ is N times the distance between one of the marginals. Therefore,

$$\gamma(a = 1) = \frac{\mathcal{D}(p_0(U_1)||p_1(U_1))}{\mathcal{D}(p_0(Y_1)||p_1(Y_1))}. \quad (8)$$

We note here that γ does not change with N . This means an increase in N will *not* effect the performance of the parallel system. γ , indeed, remains constant. So, there is no gain in discrimination information as the number of detectors increase. This is in contrast to (6) which clearly depends upon N . As the example below demonstrates, increasing the number of detectors in the hybrid system can result in gains in terms of γ .

A natural question to ask now is whether there always exists an a such that the hybrid architecture outperforms the parallel architecture? We, unfortunately, were unable to prove such an a always exists; however, simply requiring (8) to be less than (6) leads to the following inequality.

$$\mathcal{D}(p_0(U_1)||p_1(U_1)) < \mathcal{D}(p_0(U_n|U_{n-1})||p_1(U_n|U_{n-1}))$$

As one would expect, this means the interactions between the detectors must somehow produce dependencies between $p_0(\mathbf{U})$ and $p_1(\mathbf{U})$ such that the KL distance increases. Through experimentation, we found that the optimal a and the range of parameter values which produce gain, strongly depend on the signal-to-noise ratio. As the signal-to-noise ratio decreases, the value of the optimal a decreases and the range of gainful parameters shrink.

Thirdly, if the fusion center blocks the inputs of the first $(N-1)$ detectors but outputs the N^{th} detector's input, the hybrid architecture reduces to the tandem architecture. (See figure 1). Surprisingly, this simple observation, combined with a result of [8], leads to the conclusion that the hybrid architecture will always perform as least as well the tandem architecture. In [8], Lexa and Johnson showed that γ across a detector is maximized if it performs a likelihood ratio test. In the special case where the hybrid architecture reduces to the tandem architecture, the fusion center clearly does not perform a likelihood ratio test. Thus, γ across a likelihood ratio fusion center, is greater than or equal to γ across the tandem fusion center. Overall, based on the system theory outlined in [2], we have $\gamma_{\text{hybrid}} \geq \gamma_{\text{tandem}}$.

We now present an illustrative example. Let the input observations Y_n be Gaussian with mean $m_0 = 0$ (under H_0) and $m_1 = 1$ (under H_1) and variance $\sigma^2 = 1$. In this case, the likelihood ratio is given by $\Lambda(Y_n) = \exp(Y_n - \frac{1}{2})$. To simplify the problem further, we equivalently work with the sufficient statistic $\Upsilon(Y_n) = Y_n$. The constant control decision rules consequently become

$$Y_1 \underset{H_0}{\overset{H_1}{\gtrless}} \ln t + \frac{1}{2}$$

$$Y_n \underset{H_0}{\overset{H_1}{\gtrless}} \ln t a^{U_{n-1}} + \frac{1}{2}, \quad 2 \leq n \leq N$$

The baseline threshold t is set by specifying the probability of false alarm for a single detector operating in isolation. Letting $P_F = .05$, results in $t = 3.142$. The input KL distance is given by $N/2$ (see [9]). The calculation of the output joint probability distribution is straightforward and the results are plotted in the top panel of figure 2. The curves plot γ and quantify the loss in the ability to discriminate Y_1 from Y_0 for various values of N . Notice the hybrid structure produces large gains relative to the parallel architecture (relative to the flat line). This means that for certain values of a , the interactions between the local detectors effectively pushes $p_0(\mathbf{U})$ further apart from $p_1(\mathbf{U})$, making it easier to distinguish between them. Asymptotically, γ reaches its maximum at $a = 2.79$. For $N = 30$, $\mathcal{D}(p_0(\mathbf{U})||p_1(\mathbf{U}))$ approximately equals 7.3, compared to 4.63 for the parallel architecture. Using these distances in conjunction with the DPT, we can obtain lower bounds to the fusion center's minimum miss probability α^* [5]. With the fusion center's false

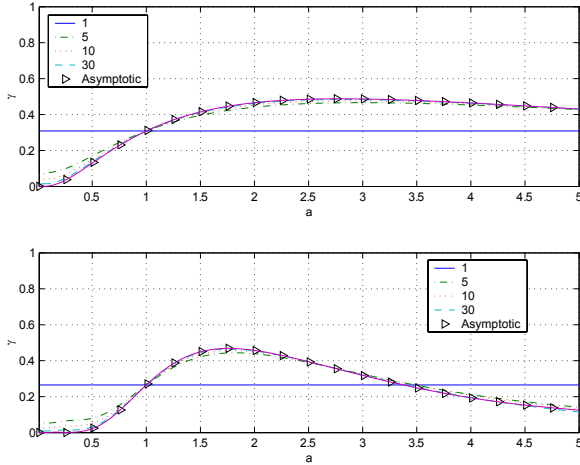


Figure 2: The information transfer ratio is plotted as a function of the linking parameter a when the input observations are iid Gaussian. The top panel shows the case when $\sigma = 1$, the bottom panel $\sigma = 2$.

alarm probability fixed at .05, $\mathcal{D}(p_0(\mathbf{U})\|p_1(\mathbf{U})) = 7.3$ implies $\alpha^* \geq .0004$. On the other hand, for the parallel system, $\mathcal{D}(p_0(\mathbf{U})\|p_1(\mathbf{U})) = 4.63$ implies $\alpha^* \geq .0062$.

The bottom plot of figure 2 plots γ when $\sigma = 2$. In this case, the gain, relative to the parallel system, is even greater than before, but the KL distance at the input of the fusion center is smaller ($\mathcal{D}(p_0(\mathbf{U})\|p_1(\mathbf{U})) = 1.741$ for $a = 1.77$).

Lastly, for both of the above cases, a γ of one reflects the performance of the optimal centralized detector because in a centralized detector the inputs \mathbf{Y} feed directly into the fusion center. Thus, the plots in figure 2 also provide the relative performance of the hybrid system compared to the optimal centralized detector.

4. CONCLUSION

The theory of information processing affords the opportunity to reexamine old distributed detection problems and provides tools to study new ones. The development of sensor networks will almost assuredly require extending current knowledge of distributed detection systems. We believe information processing provides an attractive approach to meet this demand. The hybrid system described here models a group of spatially distributed nodes all sensing the same phenomenon. If their task is signal detection, the above results suggest that even simple communication between the nodes may significantly improve performance. In this context, the fusion center is assumed to be one of the nodes within the sensor network and not a processing unit distantly located. We believe this to be a reasonable as-

sumption considering the ever increasing signal processing capabilities of sensor nodes. The ordering of the detectors in the hybrid system admittedly places additional overhead on the communication protocols but allows for simple decision rules. On the other hand, increased complexity of the decision rules could loosen the ordering constraints, producing a more practical sensor network model. For example, each node could just broadcast its decisions to neighboring nodes. It is to these types of systems we would like to apply the theory of information processing in the future.

5. REFERENCES

- [1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01)*, May 7-11 2001, vol. 4, pp. 2033–2036.
- [2] S. Sinanovic and D.H. Johnson, "Towards a theory of information processing," *Submitted to IEEE Trans. Signal Processing*, 2002.
- [3] R. Visvanathan and P. Varshney, "Distributed detection with multiple sensors: Part I-Fundamentals," *Proc. of the IEEE*, vol. 85, no. 1, pp. 54–63, Jan 1997.
- [4] J.N. Tsitsiklis, *Advances in Statistical Signal Processing*, vol. 2, chapter Decentralized Detection, pp. 297–344, JAI Press, 1993.
- [5] S. Kullback, *Information Theory and Statistics*, Wiley, New York, 1959.
- [6] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley, 1991.
- [7] Z.B. Tang, K.R. Pattipati, and D.L. Kleinman, "Optimization of detection networks: Part I-Tandem structures," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 21, no. 5, pp. 1044–1059, Sep 1991.
- [8] M.A. Lexa and D.H. Johnson, "A new look at the informational gain of soft decisions," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, April 6-10 2003, vol. 4, pp. 21–24.
- [9] D.H. Johnson and G. Orsak, "Relation of signal set choice to the performance of optimal non-gaussian detectors," *IEEE Trans. on Commun.*, vol. 41, no. 9, pp. 1319–1328, Sep. 1993.