Asymptotically Optimal Decentralized Type-Based Detection in Wireless Sensor Networks

Ke Liu and Akbar M. Sayeed
Department of Electrical and Computer Engineering
University of Wisconsin–Madison
1415 Engineering Drive, Madison, WI 53706
kliu@cae.wisc.edu, akbar@engr.wisc.edu
Corresponding author: Akbar M. Sayeed - Phone: (608)265-4731, Fax: (608)265-4623

Abstract— We propose a novel decentralized detection scheme based on the method of types and type detectors. Contrary to conventional methods, the proposed decentralized type detection (DTD) scheme asymptotically achieves the performance of centralized detection, provided there exists an arbitrarily small (but non-zero) communication capacity between the sensing nodes and the decision center. The sensing nodes extract and transmit the type information of the observed data sequence, which is then recovered and used by a type detector at the decision center. The key to DTD is that the required network information flow (type information) is of asymptotically vanishing entropy rate but still sufficient for the hypothesis testing problem. The simple structure of type sensors greatly reduces system cost in terms of computational and communication capabilities for sensor nodes. In particular, the sensor nodes need not know the signal statistics under different hypothesis. Theoretical analysis and numerical simulations demonstrate the excellent performance of the proposed DTD scheme.

I. INTRODUCTION

The *decentralized* detection or hypothesis testing problem represents a large number of applications in sensor networks, where a decision center chooses a hypothesis based on measurements at multiple (remote) sensing nodes. Typically, sensor nodes have limited computational and communication capabilities. In view of channel noise and uncertainty which corrupts transmission from sensing nodes to decision center, the conventional wisdom is that decentralized detection incurs a performance loss compared to its centralized counterpart, *centralized detection* [1], [2], which assumes perfect availability of all sensor data at the decision center.

However, this traditional view is challenged in this paper. We propose a novel, *asymptotically optimal*, decentralized detection scheme that achieves the performance of centralized detection as the length of observed data sequence increases. The key idea behind our approach is that *it is the empirical distributions rather than the actual values of observed data sequence that govern optimal hypothesis testing* [2]. For discrete random variables, the *type* of a sequence measures the its empirical distribution. It is crucial that the type information grows only polynomially with the sequence length, and thus has a vanishing entropy rate. Therefore, it can be reliably transmitted to the remote decision center provided that capacity of the channel from sensing nodes to the decision center is non-zero, no matter how small it is. The extension to

continuous alphabet can be achieved by various conventional quantization methods. In other words, we can interpret the proposed decentralized type detection scheme as a minimal compression of source data information that puts asymptotically vanishing communication burden on the network without sacrificing detection performance. Another attractive feature of the type-based detection is that the sensing node does not need the statistics information about different hypothesis— it simply acts as a dumb counter.

In fact, the use of type in statistical inference context has been investigated in the literature (see, e.g., [3] for a review) where the asymptotic optimality of type-based methods has been observed for many cases. But in this paper, we focus primarily on the effect of noisy communication channel upon the distributed detection problem. We emphasize the application of the method of types to sensor networks and its associated engineering insights for efficient information and signal processing in sensor networks.

The rest of this paper is organized as follows. The method of types and type detection are reviewed in Section II. Section III presents the proposed decentralized type detection (DTD) framework. Due to space limitation, the proof of asymptotic optimality is not included, but can be found in [4]. Section IV contains preliminary numerical results, which demonstrate the superior performance of the proposed DTD scheme compared to conventional methods.

II. THE METHOD OF TYPES AND TYPE DETECTORS

A. The Method of Types

The method of types was fully developed by Csiszár and Körner [5]. It is a powerful technique based on the large deviation theory and deals with sequences with the same empirical distribution or "type".

The following notations will be used throughout. Let $\mathcal{A}$ be the discrete alphabet with size $|\mathcal{A}| = A$. Sequence $x_1, \ldots, x_n$ may be denoted by $\mathbf{x}$, or x_1^n when dimension is emphasized. The set of all probability distributions (PD's) on $\mathcal{A}$ is denoted by $\mathcal{P}(\mathcal{A})$ or simply $\mathcal{P}$ if the alphabet $\mathcal{A}$ is clear from the context. For PD's P and Q, $H(P)$ denotes entropy and $D(P\|Q)$ denotes information divergence or the *Kullback Leibler* (KL) distance [2]: $H(P) = -\sum_{a \in \mathcal{A}} P(a) \log P(a)$

and $D(P\|Q) = \sum_{a \in \mathcal{A}} P(a) \log \frac{P(a)}{Q(a)}$. The base of $\log$ and $\exp$ is implicitly assumed to be e.

Definition 1: The type $P_{\mathbf{x}}$ (or empirical probability distribution) of a sequence $x_1, \ldots, x_n$ is the relative frequency of each symbol of $\mathcal{A}$, i.e.,

$$P_{\mathbf{x}}(a) = N(a|\mathbf{x})/n, \quad \forall a \in \mathcal{A}, \qquad (1)$$

where $N(a|\mathbf{x})$ is the number of occurrences of the symbol a in the sequence $\mathbf{x} \in \mathcal{A}^n$. Let $\mathcal{P}^n(\mathcal{A}) \subset \mathcal{P}(\mathcal{A})$ be the set of types on $\mathcal{A}^n$. For $P \in \mathcal{P}^n(\mathcal{A})$, the type class $T(P)$ denotes the set of all sequences with the same type P.

The following simple upper bound on $|\mathcal{P}^n(\mathcal{A})|$ is sufficient for most applications.

Proposition 1 (see Theorem 12.1.1 in [2]):

$$|\mathcal{P}^n(\mathcal{A})| \leq (n+1)^A. \qquad (2)$$

If $X_1, \ldots, X_n$ are drawn i.i.d. according to a PD $Q \in \mathcal{P}(\mathcal{A})$, then the probability of $\mathbf{x}$ depends only on its type $P_{\mathbf{x}}$

$$\begin{aligned} Q^n(\mathbf{x}) &= \prod_{a \in \mathcal{A}} Q(a)^{nP_{\mathbf{x}}(a)} \\ &= \exp\{-n[H(P_{\mathbf{x}}) + D(P_{\mathbf{x}}\|Q)]\}. \end{aligned} \qquad (3)$$

B. Type Detectors

The task of detection can be generally characterized as deciding the best explanation among a set of different hypotheses for the observed data. Let H_m ($1 \leq m \leq M$) denote the mth hypothesis under which the data distribution is Q_m. There are two (almost equivalent) formulations concerning the detection performance. One is the *Neyman-Pearson* formulation which considers the detection error for each hypothesis. The other is the *Bayes* formulation which weighs all the detection errors together. Given the observed data $\mathbf{x}_1^n$, the best Neyman-Pearson detector is the likelihood ratio test, where the detector will choose a hypothesis according to the ratio of various pairs of $Q_m^n(\mathbf{x})$ and $Q_l^n(\mathbf{x})$ [1]. On the other hand, the optimal Bayes detector is a MAP (the maximum a posterior probability) detector, which chooses the hypothesis with the maximum value of $\pi(m)Q_m^n(\mathbf{x})$ where $\pi(m)$ for $1 \leq m \leq M$ is the prior probability of the mth hypothesis [1]. When the prior probability for all hypotheses are equal, one has the standard *maximum likelihood* (ML) detector, which chooses the hypothesis with the largest likelihood

$$d(\mathbf{x}) = \arg \max_{1 \leq m \leq M} Q_m^n(\mathbf{x}). \qquad (4)$$

Therefore, given the observed sequence $X_1, \ldots, X_n$, the sufficient statistics for the detection problem are all the Q_m-probabilities of the given sequence $\mathbf{x}$, that is, $Q_m^n(\mathbf{x})$ for $1 \leq m \leq M$.

Assuming X_i's to be i.i.d., the Q_m-probability of $\mathbf{x}$ depends only on the type P_x by (3). It follows that two observed sequences with the same type will yield identical decision statistics, thus indistinguishable from the viewpoint of optimal detection. In other words, it is the type of the observed sequence, not its particular value, that determines the operation of the optimal detector. Thus, we introduce the concept of a *type detector*.

Definition 2: A type detector is one that operates on the type of the observed data.

Optimal detectors such as likelihood ratio test, MAP (or ML) detector are particular examples of type detectors, which implies the optimality of type detector. Classical optimal detection can be alternatively interpreted in the framework of type detection as first extracting the type information from the observed sequence and then making a decision accordingly. Mathematically, let $S_1^n, \ldots, S_M^n$ be a partition of the type space $\mathcal{P}^n$ with S_m associated with hypothesis m, that is,

$$S_i^n \cap S_j^n = \phi, \ \forall i \neq j, \ \text{and} \ \bigcup_i S_i^n = \mathcal{P}^n. \qquad (5)$$

The detection is given by

$$d(\mathbf{x}) = \arg_{1 \leq m \leq M}\{P_{\mathbf{x}} \in S_m^n \text{is true}\}. \qquad (6)$$
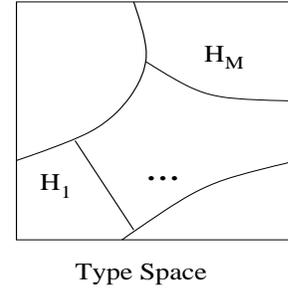

Fig. 1. A schematic illustrating the partition of the type space into disjoint subsets corresponding to different hypothesis.

Fig. 1 illustrates the interpretation of detection via types. The type space $\mathcal{P}^n$, dense in the simplex $\mathcal{P}$, is partitioned into M decision regions, each assigned to a hypothesis. A particular hypothesis is chosen by the detector if the observed type belongs to the corresponding decision region. In the framework of types, it is readily verified that ML detection amounts to

$$d(\mathbf{x}) = \arg \min_{1 \leq m \leq M} D(P_{\mathbf{x}}\|Q_m), \qquad (7)$$

that is, the detector chooses the mth hypothesis if the type (or empirical distribution) of the data is closest to Q_m in KL-distance. The type decision region in this case is given by

$$S_m^{ML,n} = \{P : D(P\|Q_m) < D(P\|Q_l), \forall l \neq m\}. \qquad (8)$$

III. DECENTRALIZED TYPE DETECTION IN SENSOR NETWORKS

A. Asymptotic Optimality of Decentralized Detection

Consider K independent sensors. Let $X_1^{(k)}, \ldots, X_n^{(k)}$ be the observed i.i.d. sequence (of length n) at the kth sensor. Denote by C the minimum communication capacity of each sensor to the decision center. The total channel use, t, by remote sensor nodes is required to be on the order of the sequence length n, that is,

$$\frac{t}{n} \sim O(1), \quad \text{as } n \to 0. \qquad (9)$$

Given a detection problem involving M different hypotheses with PD's $Q_1, \ldots, Q_M$, the ideal centralized detection has all sensor data available at the decision center. Let E_c be the optimal exponent of centralized detection on K sensors for hypothesis m, that is,

$$\lim_{n \to \infty} -\frac{1}{nK} \log \alpha_m^{nK} = E_c \qquad (10)$$

where α is the associated detection error probability. Let E_d denote the error exponent for decentralized type detection scheme.

Theorem 1 (Asymptotic Optimality): Assume $D = \min_{l \neq m} D(Q_l \| Q_m) < \infty$. If $C > 0$, then

$$(E_d)_{\text{opt}} = E_c, \qquad (11)$$

that is, the decentralized type detection achieves the same optimal performance as its centralized counterpart, provided the communication capacity between sensor nodes and decision center is strictly positive (but maybe arbitrary small). Moreover, the optimal decentralized detection is achieved by transporting type information from the sensing nodes and using type detection at the decision center.
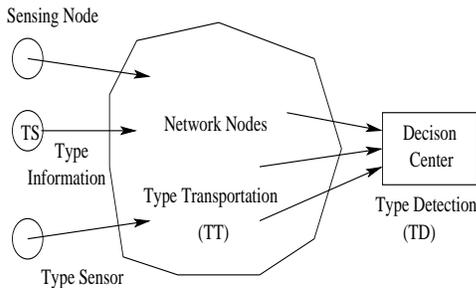
B. Decentralized Type Detection

Fig. 2. A schematic illustrating the DTD framework. The remote sensing nodes are type sensors (TS) that collect the type information of observed data sequence. Type information from various sensing nodes is transported to the decision center, where type detection (TD) is used to make final detection.

The proposed framework of decentralized type detection in sensor networks is illustrated in Fig. 2. For brevity, we assume discrete random variables (or fixed quantization level for continuous alphabets). The sensor nodes are configured as *type sensors* which simply record the type of the observed sequence. Such an operation involves nothing more than counting of the relative frequency for each symbol in the alphabet. Upon observing a sequence of data, each type sensor codes its message according to the type it has observed and transmits the type information to the remote decision center, which we call the *type transportation*. The type transportation from sensor nodes to decision center may involve the collaboration of other intermediate nodes depending on network communication protocols. The decision center receives the type information from all remote sensors. It then combines received type information to compute the *joint type* on all

sensor data, based on which the type detector makes the final decision.

There are several major advantages of the DTD framework. First, the sensor nodes can be both "cheap" in the sense that only counting is sufficient, and "dumb" in the sense that the sensor nodes need no knowledge of the detection task intended in the decision center. The decision center can exert a great level of flexibility in its detection tasks without updating the remote sensor. This feature is particularly helpful to reducing the system cost. In contrast, existing methods either demand sophisticated decision making at sensor nodes or require knowledge of signal statistics for off-line configuration of data processing at the sensor nodes [6], [7]. Second, the type information is of asymptotically vanishing (entropy) rate, compared with existing methods that have a non-trivial data flow. Thus, the proposed framework puts an asymptotically vanishing communication burden on the network, which, in turn, affords great flexibility in designing communication protocols.
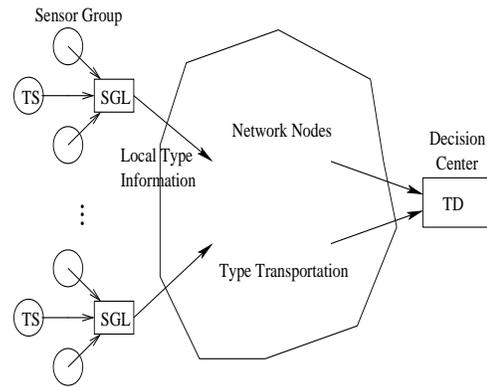

Fig. 3. Remote sensing nodes are arranged into small local groups among which exchange of data at relatively high speed is feasible. Each group selects a group leader (SGL), which computes the joint type of all the sensor in its group and then transmits its type information to the remote decision center.

However, the efficiency of the DTD framework is critically controlled by the length of observed sequence at sensor nodes. The longer the sequence, the smaller the relative sensor information flow. If sensor nodes can only observe a very short sequence (one data record, for instance), one can divide sensor nodes in local groups. Relatively higher communication can be attained within a group; close sensor nodes enjoy larger link capacity than distant nodes. The collaboration is carried out in a group such that the group leader in a group can obtain the joint type of all observed data in the group. From this point on, individual group leaders relay type information to the remote decision center. Such a modified scheme is illustrated in Fig. 3. Since the information flow from sensor group leaders to decision center can be made small, the link capacity requirement on decision center is thus significantly reduced. For example, the decision center can now be far from the sensor deployment without severely handicapping its detection performance.

IV. NUMERICAL RESULTS

In this section, we present numerical results to illustrate the performance of decentralized type detection. Source data are i.i.d. Bernoulli random variables (with symbol "1" and "0") with parameters corresponding to two different hypotheses. To simplify the simulation, a dedicated additive white Gaussian noise (AWGN) channel is assumed between each sensing node and decision center. The following detection schemes are considered:

1) *Centralized Detection*: Data sequence observed by all sensing nodes are perfectly available at the decision center. The performance of centralized detection upper-bounds that of any decentralized detection scheme.

2) *Decentralized Type Detection*: Sensing node prepares the type information according its observation data. A simple repetition code is used for channel coding and channel modulation is chosen to be Binary Phase Shift Keying (BPSK). To illustrate the operation at the sensing nodes, we consider, for example, that a sequence $1100, 0010, 0000, 110$ of length 15 is observed by a sensing node. It counts the number of occurrence of "1" and "0" to get the type information $N(1) = 5$ and $N(0) = 10$ in this case, whose binary representation is $0101, 1010$, a total of 8 bits. Since we allow transmission duration to be about the same length as sequence length, the type information is repeated until it fills all the transmission time. In this case, the information packet sent through the channel is, $00, 11, 00, 11, 11, 00, 11, 00$, noting that each bit is repeated twice, that is, a repetition code of rate $1/2$ is used to encode the type information. At the decision center, received signals from sensing nodes are demodulated and decoded to reconstruct the type information packet of each sensing node. For a successful transmission, type information $N'(1) = 5$ and $N'(0) = 10$ is recovered at the decision center. Otherwise, a detection error may happen due to erroneous type information. However, the probability of such an error event decays exponentially as the length of data sequence increases.

3) *Soft Data Fusion*: The BPSK modulation is directly used to send sensor data to the decision center, which induces an overall distribution on the received signal at the decision center. Such a composite distribution is due to both source distribution and channel noise. For Bernoulli parameter p, the composite probability density function is given by

$$f(x) = pG(x - 1) + (1 - p)G(x + 1) \qquad (12)$$

where $G(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-x^2/2\sigma^2}$ with σ specified by channel SNR. The ML detection on the composite distribution is carried out at the decision center.

4) *Hard Data Fusion*: Similar to the soft data fusion method, sensor data are sent to decision center by BPSK. But, instead of detection based on composite distribution, the decision center first decodes sensor data,

which is a sequence of "1" and "0". Then, ML detection is performed as if the reconstructed data sequences are perfect without transmission error.

We emphasize that the channel SNR is kept the same for all decentralized detection schemes to keep the comparison fair.
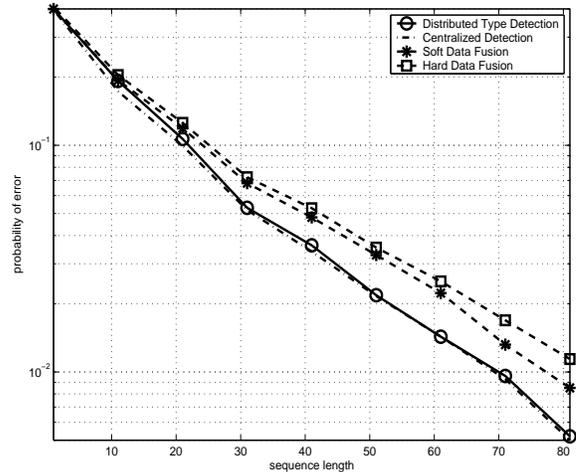

Fig. 4. Detection error probability as a function of data sequence length. The channel SNR is 0dB. Two hypotheses on source are $p_0 = 0.6$ or $p_1 = 0.4$. The number of sensors is $K = 2$.

Fig. 4 plots a performance comparison between aforementioned detection schemes. The channel is very noisy (0dB), which suggests a small transmission capacity. It is clear that detection error probability is in exponential decay with the sequence length. However, the decentralized type detection exhibits almost identical performance as that of centralized detection, while performance loss associated with data fusion schemes is also evident.

V. CONCLUSIONS

We have presented a novel decentralized detection based on the method of types that is asymptotically optimal. Our results demonstrate great potential of the proposed DTD scheme to reduce the communication burden on the network. In particular, the DTD framework can be applied to the MAC channel situation where all sensor nodes share the same channel.

REFERENCES

[1] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.
[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[3] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2300–2324, Oct. 1998.
[4] K. Liu and A. M. Sayeed, "On asymptotic optimality of decentralized detection in wireless sensor networks," Preprint available at http://dune.ece.wisc.edu, Oct. 2003.
[5] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[6] J. N. Tsitsiklis, "Decentralized detection," *Advances in Statistical Signal Processing*, vol. 2, pp. 297–344, 1993.
[7] M. Longo, T. D. Lookabaugh, and R. M. Gray, "Quantization for decentralized hypothesis testing under communication constraints," *IEEE Trans. Inform. Theory*, vol. 36, no. 2, pp. 241–255, Mar. 1990.