Optimal Remote Estimation of Discrete Random Variables Over the Collision Channel

Marcos M. Vasconcelos and Nuno C. Martins

***Abstract*—Consider that a remote estimator receives data from multiple sensors via an interference-limited wireless network modeled by a collision channel. Each sensor assesses a discrete random variable and must decide whether to attempt to transmit it to the remote estimator, or remain silent. The random variables are independent and there is no communication among the sensors, which precludes coordinated transmission strategies. The estimator seeks to estimate all the random variables. The collision channel outputs a collision symbol when two or more sensors attempt transmission, and it acts as an ideal link otherwise. The goal is to design transmission policies that are globally optimal with respect to two costs. The first cost is a convex combination of the probabilities that the estimates of the individual random variables are incorrect, and the second is the probability that the estimates of one or more random variables are incorrect. We show that the problems defined by these costs admit simple deterministic globally optimal structures within which numerical optimization is tractable, even when the support of the random variables is infinite.**

***Index Terms*—Decentralized estimation, optimization, maximum a posteriori estimator, team decision theory, networked estimation, concave minimization problems.**

I. Introduction

Over the last few years, Cyber-Physical systems have emerged as a framework of system design where multiple agents sense, communicate over a network and actuate on a physical system, operating as a team to achieve a common goal [1]. When the objective is to optimize a certain cost, the system designer's task is to solve a problem of decentralized decision making, where the agents have access to different information and choose an action that incurs in a cost that depends on the actions of all the decision makers. Furthermore, network constrains and stringent delay requirements on the flow of information between the agents, forces them to make efficient use of the communication resources.

We consider the following Bayesian estimation problem illustrated by the block diagram of Figure 1. Two sensors observing independent discrete random variables, decide whether to communicate their measurements to a remote estimator over a collision channel according to transmission policies. The communication constraint imposed by the collision channel is such that only one sensor can transmit its measurement perfectly; and if more than one sensor transmit simultaneously, a collision is declared. Upon observing the channel output, the estimator forms estimates of all the measured random variables. Our goal is to find transmission policies that optimize two performance criteria involving probabilities of estimation error.

M. M. Vasconcelos is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA. E-mail: mvasconc@usc.edu. N. C. Martins is with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD, 20742 USA. E-mail: nmartins@umd.edu.

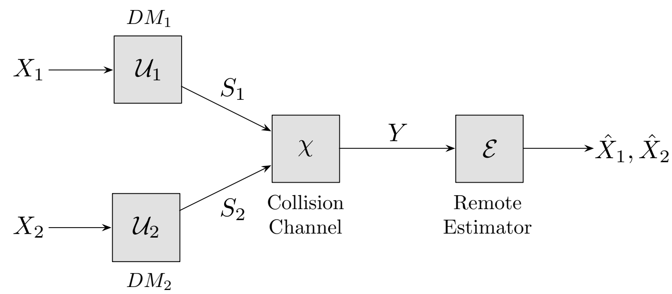

Fig. 1. Schematic representation of distributed estimation over a collision channel.

A. Applications

The collision channel captures data-transfer restrictions that may result, for instance, from the interference caused by wireless transmitters that share the same frequency band and are not capable of executing scheduling or random access protocols. These constraints are present in large networks of simple devices, such as tiny low-power sensors. Examples include nanoscale intra-body networks for health monitoring and drug delivery systems; networks for environmental monitoring of air pollution, water quality and biodiversity control [2], [3]. Remote estimation systems of this type can also be applied in scenarios where the devices are heterogeneous and there is a strict requirement for real-time wireless networking. Collisions may also occur in ad-hoc networks that lack a coordination protocol among the devices [4]; data centers, which are subject to cascading power failures [5] or cyber-attacks [6] that must be detected in minimal time and as accurately as possible.

B. Related literature, prior work and contributions

Problems of distributed decision making such as the one in this paper fall into the category of team decision problems with discrete observation and action spaces, and a nonclassical information structure [7]. It is known that problems in this class are in general NP-complete [8]. One possible approach to solve team problems of this type is to use an approximation technique, which guarantees that a suboptimal strategy is within a fixed bound of the globally optimal one [9]. Another set of results pertains to a class of problems for which the cost satisfies a property known as multimodularity, which allows the characterization of the set of person-by-person optimal solutions and efficient algorithms for searching for a globally optimal solution [10]. Despite of the inherent difficulty of such discrete team decision problems, the remote estimation problem formulated in this paper admits a characterization of the person-by-person optimal solutions with structural results that either significantly reduce the search space for globally optimal solutions or solve it completely, depending on which of the two criteria is used as objective. Our results are exact and do not make use of any approximation techniques.

There exists a large literature on estimation and control of dynamical systems over wireless networks with a continuous state space. Even though most of modern systems are either digital or hybrid in nature, few papers deal with the case when the dynamical system has a discrete state space, e.g. [11]. A class of problems in information theory known as real-time coding and decoding of discrete Markov sources is equivalent to remote estimation problems of plants with a discrete state space. Notably, many contributions in this area are derived using ideas of stochastic control such as identifying sufficient statistics for the optimal coding and decoding policies. Structural results of optimal policies for sequential problem formulations involving a single sensor and a remote estimator were obtained in [12]–[14] and [15]. The problem of estimating an independent and identically distributed discrete source observed by a single sensor with a limited number of measurements and a probability of error criterion was solved in [16]. The case in which the state of a linear time invariant system, driven by white noise, is estimated over a costly channel is addressed in [17]. A sequential, multi-sensor, real-time communication problem over parallel channels was investigated by [18].

Our problem is an instance of a one-shot, remote estimation (real-time communication) problem over a multiple access channel with independent discrete observations. The problem of estimating independent continuous random variables over a collision channel while minimizing a mean squared error criterion was solved in [19], where it was shown the existence of globally optimal deterministic transmission strategies with a threshold structure. The authors of [19] also show that this result is independent of the probability distributions of the observed random variables. Here, we solve two related problems where the objective is to optimize a probability of error criterion, which is a popular metric used in estimation of discrete random variables, statistical hypothesis testing and rate distortion theory [20]. A related team decision problem involving the optimization of a probability of error criterion similar to the one considered here was studied in [21], where the optimal strategy is based on ideas from coding theory. The main results of this paper is the discovery of a deterministic structure of certain globally optimal solutions for any pair of probability mass functions that characterize the observed data. An interesting feature of our results is that the constraint imposed by the channel results in optimal policies for which only certain subsets of the measurements are transmitted and others are not. This phenomenon is known as data reduction via *censoring* and was explored for continuous observations with a mean squared error criterion in [22].

Although our problem is based on a framework introduced in [19], the techniques and results shown here do not follow from our previous work and have not appeared elsewhere in the literature.

C. Paper organization

The paper is structured in six sections, including the Introduction. In Section II, we describe the problem setup; define the two fidelity criteria used to obtain optimal policies: the aggregate probability of estimation error and the total probability of estimation error; and state our main structural results. In Section III, we prove the structural result for the aggregate probability of error and in Section IV, we prove the result for the total probability of error. In Section V, we discuss the extension of these results to teams of N decision makers. The paper ends in Section VI with our conclusions and suggestions for future work.

D. Notation

We adopt the following notation: Functions and functionals are denoted using calligraphic letters such as $\mathcal{F}$ or $\mathscr{F}$. Sets are represented in blackboard bold font, such as $\mathbb{A}$. The cardinality of a set $\mathbb{A}$ is denoted by $|\mathbb{A}|$. The set of real numbers is denoted by $\mathbb{R}$. If $\mathbb{A}$ is a subset of $\mathbb{B}$ then $\mathbb{B}\backslash\mathbb{A}$ represents the set of elements in $\mathbb{B}$ that are not in $\mathbb{A}$. Discrete random variables, vectors of discrete random variables and discrete general random elements are represented using upper case letters, such as W. Realizations of W are represented by the corresponding lower case letter w. The probability of an event $\mathfrak{E}$ is denoted by $\mathbf{P}(\mathfrak{E})$. The probability mass function of a Bernoulli random variable W, for which $\mathbb{P}(W=1)=\delta$, is denoted as $\mathcal{B}(\delta)$.

We also adopt the following conventions:

- If $\mathfrak{A}$ and $\mathfrak{B}$ are two events for which $\mathbf{P}(\mathfrak{B})=0$ then we adopt the convention that $\mathbf{P}(\mathfrak{A}|\mathfrak{B})=0$.
- Consider that a subset $\mathbb{W}$ of $\mathbb{R}^n$ and a function $\mathcal{F}:\mathbb{W}\rightarrow\mathbb{R}$ are given. If $\overline{\mathbb{W}}$ is the subset of elements that maximize $\mathcal{F}$ then $\arg\max_{\alpha\in\mathbb{W}}\mathcal{F}(\alpha)$ is the greatest number in $\overline{\mathbb{W}}$ according to the lexicographic order.

II. General problem setup

Consider two independent discrete random variables X_1 and X_2 taking values on finite or countably infinite alphabets $\mathbb{X}_1$ and $\mathbb{X}_2$, respectively. Each random variable X_i is distributed according to a given probability mass function $p_{X_i}(x_i)$ on $\mathbb{X}_i$, $i\in\{1,2\}$. Without loss of generality, we assume that every element of $\mathbb{X}_i$ occurs with a strictly positive probability, for $i\in\{1,2\}$.

There are two sensors[1] denoted as DM_1 and DM_2 that measure X_1 and X_2, respectively. Each decision maker DM_i observes a realization of X_i, and must decide whether to remain silent or attempt to transmit x_i to the estimator. The decision to attempt a transmission or not is represented by a binary random variable $U_i\in\{0,1\}$, where $U_i=1$ denotes the decision to attempt a transmission and with $U_i=0$ DM_i remains silent. The decision by DM_i on whether to

[1] We use the terminology *sensor* and *decision maker* (DM) interchangeably throughout the paper.

transmit is based solely on its measurement x_i, according to a transmission policy $\mathcal{U}_i$ defined as follows:

Definition 1 (Transmission Policies): The transmission policy for DM_i is specified by a function $\mathcal{U}_i : \mathbb{X}_i \to [0,1]$ that governs a randomized strategy as follows:

$$\mathbf{P}(U_i = 1 | X_i = x_i) \overset{\text{def}}{=} \mathcal{U}_i(x_i), \quad i \in \{1,2\}. \tag{1}$$

The set of all transmission policies for DM_i is denoted by $\mathbb{U}_i \overset{\text{def}}{=} [0,1]^{|\mathbb{X}_i|}$.

Assumption 1: We assume that the randomization, which generates U_1 and U_2, according to Eq. (1), is such that the pairs (U_1, X_1) and (U_2, X_2) are independent:

$$\begin{aligned}\mathbf{P}(U_1 = \mu_1, U_2 = \mu_2, X_1 = \alpha_1, X_2 = \alpha_2) = \\ \mathbf{P}(U_1 = \mu_1, X_1 = \alpha_1)\mathbf{P}(U_2 = \mu_2, X_2 = \alpha_2).\end{aligned} \tag{2}$$

Definition 2: The measurement X_i and the decision U_i by DM_i specify the random element S_i, which will be used as a channel input, as follows:

$$s_i \overset{\text{def}}{=} \begin{cases} (i, x_i) & \text{if } u_i = 1 \\ \varnothing & \text{if } u_i = 0 \end{cases}, \quad i \in \{1,2\}, \tag{3}$$

Each random element S_i takes values in $\{\varnothing\} \cup \{(i, \alpha_i) \mid \alpha_i \in \mathbb{X}_i\}$, where the symbol $\varnothing$ denotes *no-transmission*.

Remark 1: Notice that S_1 and S_2 contain the identification number of its sender. This allows the estimator to determine unambiguously the origin of every successful transmission.

Definition 3 (Collision Channel): The collision channel takes S_1 and S_2 as inputs. The output Y of the collision channel is characterized by the following deterministic map:

$$y = \chi(s_1, s_2) \overset{\text{def}}{=} \begin{cases} s_1 & \text{if } s_1 \neq \varnothing, \ s_2 = \varnothing \\ s_2 & \text{if } s_1 = \varnothing, \ s_2 \neq \varnothing \\ \varnothing & \text{if } s_1 = \varnothing, \ s_2 = \varnothing \\ \mathfrak{C} & \text{if } s_1 \neq \varnothing, \ s_2 \neq \varnothing, \end{cases} \tag{4}$$

where the symbol $\mathfrak{C}$ denotes a *collision*.

Remark 2: The fact that the collision channel discerns between a collision $\mathfrak{C}$ and the absence of a transmission (indicated by $\varnothing$), makes it fundamentally different from the erasure link commonly found in the literature of remote control and estimation, such as in [23]. This creates an opportunity to improve estimation performance by implicitly encoding information in $\mathfrak{C}$ and $\varnothing$.

A. The aggregate probability of estimation error criterion

For any given policy pair $(\mathcal{U}_1, \mathcal{U}_2) \in \mathbb{U}_1 \times \mathbb{U}_2$, we start by considering the following fidelity criterion consisting of a convex combination of the individual probabilities of error of estimating X_1 and X_2.

Definition 4: Define $\mathcal{J}_A : \mathbb{U}_1 \times \mathbb{U}_2 \to \mathbb{R}$ such that

$$\mathcal{J}_A(\mathcal{U}_1, \mathcal{U}_2) \overset{\text{def}}{=} \eta_1 \mathbf{P}(X_1 \neq \hat{X}_1) + \eta_2 \mathbf{P}(X_2 \neq \hat{X}_2), \tag{5}$$

where $\eta_1, \eta_2 > 0$ are given positive constants satisfying $\eta_1 + \eta_2 = 1$.

The designer can choose η_1 and η_2 to set the relative priority of each of the random variables it is interested in. It is straightforward to show that for any two transmission policies, the receiver that minimizes the cost in Eq. (5) forms a *maximum a posteriori probability* (MAP) estimate of the random variable X_i given the observed channel output Y according to functions $\mathcal{E}_i : \mathbb{Y} \to \mathbb{X}_i$ defined as

$$\mathcal{E}_i(y) \overset{\text{def}}{=} \arg\max_{\alpha \in \mathbb{X}_i} \mathbf{P}(X_i = \alpha | Y = y), \ i \in \{1,2\}. \tag{6}$$

Problem 1: Given a pair of probability mass functions p_{X_1} and p_{X_2}, find a policy pair $(\mathcal{U}_1, \mathcal{U}_2) \in \mathbb{U}_1 \times \mathbb{U}_2$ that minimizes $\mathcal{J}_A(\mathcal{U}_1, \mathcal{U}_2)$ in Eq. (5), subject to the communication constraint imposed by the collision channel of Eq. (4) and that the estimator employs the MAP rule of Eq. (6).

B. The total probability of estimation error criterion

Given a policy pair $(\mathcal{U}_1, \mathcal{U}_2) \in \mathbb{U}_1 \times \mathbb{U}_2$, we also consider the cost $\mathcal{J}_B : \mathbb{U}_1 \times \mathbb{U}_2 \to \mathbb{R}$ defined as:

$$\mathcal{J}_B(\mathcal{U}_1, \mathcal{U}_2) \overset{\text{def}}{=} \mathbf{P}(\{X_1 \neq \hat{X}_1\} \cup \{X_2 \neq \hat{X}_2\}) \tag{7}$$

which accounts for the probability that at least one estimate is incorrect.

In this case, for any two transmission policies, the receiver that minimizes the cost in Eq. (7) forms a MAP estimate of the random variables (X_1, X_2) given the observed channel output Y according to a function $\mathcal{E} : \mathbb{Y} \to \mathbb{X}_1 \times \mathbb{X}_2$ defined as follows:

$$\mathcal{E}(y) \overset{\text{def}}{=} \arg \max_{(\alpha_1, \alpha_2) \in \mathbb{X}_1 \times \mathbb{X}_2} \mathbf{P}(X_1 = \alpha_1, X_2 = \alpha_2, Y = y). \tag{8}$$

Problem 2: Given a pair of probability mass functions p_{X_1} and p_{X_2}, find a pair of policies $(\mathcal{U}_1, \mathcal{U}_2) \in \mathbb{U}_1 \times \mathbb{U}_2$ that jointly minimizes $\mathcal{J}_B(\mathcal{U}_1, \mathcal{U}_2)$ in Eq. (7), subject to the communication constraint imposed by the collision channel of Eq. (4) and that the estimator employs the MAP rule of Eq. (8).

Definition 5 (Global optimality): A pair of transimission policies $(\mathcal{U}_1^\star, \mathcal{U}_2^\star) \in \mathbb{U}_1 \times \mathbb{U}_2$ is *globally optimal* for the cost $\mathcal{J}(\mathcal{U}_1, \mathcal{U}_2)$ if the following holds:

$$\mathcal{J}\left(\mathcal{U}_1^\star, \mathcal{U}_2^\star\right) \leq \mathcal{J}\left(\mathcal{U}_1, \mathcal{U}_2\right), \quad (\mathcal{U}_1, \mathcal{U}_2) \in \mathbb{U}_1 \times \mathbb{U}_2. \tag{9}$$

C. A motivating example

Sensor scheduling is one way to guarantee that collisions never occur. However, in general, scheduling is not optimal. In order to illustrate this, consider the simple scenario where X_1 and X_2 are independent Bernoulli random variables with nondegenerate probability mass functions p_{X_1} and p_{X_2}, respectively. Using a sensor scheduling policy where only one sensor is allowed to access the channel in Problem 1, the best possible performance is given by

$$\mathcal{J}_A^{\text{sch}} = 1 - \max_{i \in \{1,2\}} \max_{x \in \{0,1\}} \eta_i p_{X_i}(x) > 0. \tag{10}$$

However, it is possible to achieve zero aggregate probability of error, for any independent Bernoulli X_1 and X_2, by using the following pair of deterministic policies $(\mathcal{U}_1^\star, \mathcal{U}_2^\star)$:

$$\mathcal{U}_i^\star(x_i) = x_i, \ i \in \{1,2\}. \tag{11}$$

The pair $(\mathcal{U}_1^\star, \mathcal{U}_2^\star)$ achieves zero aggregate probability of error because U_1 and U_2, or equivalently in this case X_1 and X_2, can be recovered from the channel output Y. This globally optimal pair of policies makes use of the distinction that the channel in Eq. (4) makes between no-transmissions and collisions to convey information about the observations to the remote estimator.

Motivated by this example, we are interested in investigating whether *there are similar strategies that are globally optimal for any given p_{X_1} and p_{X_2}.*

D. Person-by-person optimality

Problems with a non-classical information pattern, such as the ones considered here, are non-convex in general. Therefore, determining globally optimal solutions is often intractable. In Sections III and IV, we proceed to showing that there are globally optimal solutions for Problems 1 and 2 with a convenient structure, for which numerical optimization is tractable. This is accomplished via the following concept of *person-by-person* optimality [7], [10].

Definition 6 (Person-by-person optimality): A policy pair $(\mathcal{U}_1^\star,\mathcal{U}_2^\star) \in \mathbb{U}_1 \times \mathbb{U}_2$ is said to satisfy the person-by-person necessary conditions of optimality for the cost $\mathcal{J}(\mathcal{U}_1,\mathcal{U}_2)$ if the following holds:

$$\begin{aligned} \mathcal{J}(\mathcal{U}_1^\star,\mathcal{U}_2^\star) &\leq \mathcal{J}(\mathcal{U}_1,\mathcal{U}_2^\star), \quad \mathcal{U}_1 \in \mathbb{U}_1 \\ \mathcal{J}(\mathcal{U}_1^\star,\mathcal{U}_2^\star) &\leq \mathcal{J}(\mathcal{U}_1^\star,\mathcal{U}_2), \quad \mathcal{U}_2 \in \mathbb{U}_2. \end{aligned} \tag{12}$$

A policy pair that satisfies Eq. (12) is also called person-by-person optimal.

E. Main structural results

The main results of this paper pertain to unveiling the structure of globally optimal transmission policies for the problems stated in Sections II-A and II-B. One important feature of the results below is that they are independent of the distributions of the observations, and are valid even when the alphabets are countably infinite.

Before we continue, we proceed to defining a few important policies that will be used to characterize certain globally optimal solutions.

Definition 7 (An useful total order): For every discrete random variable W, taking values in $\mathbb{W}$, we define a totally ordered set $[\mathbb{W}, p_W]$ in which, for any w and $\tilde{w}$ in $[\mathbb{W}, p_W]$, $w \prec \tilde{w}$ holds when $p_W(w) < p_W(\tilde{w})$ or, for the case in which $p_W(w) = p_W(\tilde{w})$, when $w < \tilde{w}$. The elements of $[\mathbb{W}, p_W]$ are enumerated so that $w_{[1]}$ is the maximal element and $w_{[i+1]} \prec w_{[i]}$, for every integer i greater than or equal to 1. We also adopt q_W to denote the following probability mass function:

$$q_W(i) \overset{\text{def}}{=} \mathbf{P}(W = w_{[i]}), \quad 1 \leq i \leq |\mathbb{W}| \tag{13}$$

Definition 8 (Candidate optimal policies): Given a discrete random variable W, we use the total order in $[\mathbb{W}, p_W]$ to define:

$$\mathcal{V}_W^{\bar{1}}(\alpha) \overset{\text{def}}{=} \begin{cases} 0 & \text{if } \alpha = w_{[1]} \\ 1 & \text{otherwise} \end{cases}, \quad \alpha \in \mathbb{W} \tag{14}$$

$$\mathcal{V}_W^{2}(\alpha) \overset{\text{def}}{=} \begin{cases} 1 & \text{if } \alpha = w_{[2]} \\ 0 & \text{otherwise} \end{cases}, \quad \alpha \in \mathbb{W} \tag{15}$$

$$\mathcal{V}_W^{\varnothing}(\alpha) \overset{\text{def}}{=} 0, \quad \alpha \in \mathbb{W} \tag{16}$$

Theorem 1 (Globally optimal solutions for Problem 1): There exists a globally optimal solution pair $(\check{\mathcal{U}}_1,\check{\mathcal{U}}_2)$ for which each component $\check{\mathcal{U}}_i$ is one of the three policies in $\{\mathcal{V}_{X_i}^{\bar{1}}, \mathcal{V}_{X_i}^{2}, \mathcal{V}_{X_i}^{\varnothing}\}$, for $i \in \{1, 2\}$.

Remark 3: Theorem 1 implies that, regardless of the cardinality of $\mathbb{X}_1$ and $\mathbb{X}_2$, one can determine a globally optimal solution by checking *at most nine candidate solutions*. In fact, Corollary 1 shows in Section III-C that one can reduce the search to at most five candidate solutions.

Theorem 2 (Globally optimal solutions for Problem 2): The policy pair $(\mathcal{V}_{X_1}^{\bar{1}}, \mathcal{V}_{X_2}^{\bar{1}})$ is globally optimal.

Using the person-by-person optimality approach, Sections III-C and IV-B present proofs for Theorems 1 and 2, respectively.

III. Solution to Problem 1

The person-by-person approach to proving Theorem 1 involves the analysis of the associated problem of optimizing the policy of a single sensor, while keeping the policy of the other sensor fixed. This problem is precisely formulated and solved in Section III-B, where we also show that it is a concave minimization problem. Such problems are, in general, intractable, but we were able to find solutions using a two-step approach. More specifically, we obtain a lower bound that holds for any feasible policy (the converse part) and then we provide a structured deterministic policy that achieves the lower bound (the achievability part).

We start by using Bayes' rule to rewrite the cost in Eq. (7) in a way that clarifies the effect of modifying the policy of a single sensor. More specifically, from the perspective of DM_i, $i \in \{1, 2\}$, assuming that the policy used by DM_j is fixed to $\tilde{\mathcal{U}}_j \in \mathbb{U}_j$, $j \neq i$, we have:

$$\mathcal{J}_A(\mathcal{U}_i,\tilde{\mathcal{U}}_j) = \eta_i \mathbf{P}(X_i \neq \hat{X}_i) + \eta_j(\rho_{\tilde{\mathcal{U}}_j} \mathbf{P}(U_i = 1) + \theta_{\tilde{\mathcal{U}}_j}), \tag{17}$$

where

$$\rho_{\tilde{\mathcal{U}}_j} \overset{\text{def}}{=} \mathbf{P}(X_j \neq \hat{X}_j | U_i = 1) - \mathbf{P}(X_j \neq \hat{X}_j | U_i = 0) \tag{18}$$

and

$$\theta_{\tilde{\mathcal{U}}_j} \overset{\text{def}}{=} \mathbf{P}(X_j \neq \hat{X}_j | U_i = 0). \tag{19}$$

The terms $\rho_{\tilde{\mathcal{U}}_j}$ and $\theta_{\tilde{\mathcal{U}}_j}$ are constant in $\mathcal{U}_i$. In particular, $\eta_j \rho_{\tilde{\mathcal{U}}_j}$ can be interpreted as a *communication cost* incurred by DM_i when it attempts to transmit. A similar interpretation has been used in [19] and relates this problem to the multi-stage estimation case with limited actions solved in [16].

A. The communication cost and off-set terms

We proceed to characterizing the communication cost Eq. (17) and the offset terms in further detail: first by showing that they are constant in $\mathcal{U}_i$ and then establishing that they are non-negative and upper bounded by 1. These facts will be subsequently used in the proof of Theorem 1.

Proposition 1: If X_1 and X_2 are independent and Assumption 1 holds then the following inequalities are satisfied:

$$0 \leq \rho_{\tilde{\mathcal{U}}_j} \leq 1 \tag{20}$$

$$0 \leq \theta_{\tilde{\mathcal{U}}_j} \leq 1 \tag{21}$$

Proof: First, we need to show that, for $i, j \in \{1, 2\}$ and $j \neq i$, the following holds:

$$\mathcal{E}_j\left((i, x_i)\right) = \mathcal{E}_j(\varnothing),\ x_i \in \mathbb{X}_i, \tag{22}$$

which implies that for the purpose of estimating X_j, observing $Y = (i, x_i)$ at the fusion center is equivalent to receiving $Y = \varnothing$.

From the definition of the MAP estimator in Eq. (6), we have:

$$\mathcal{E}_j(\varnothing) = \arg\max_{\alpha \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha, Y = \varnothing) \tag{23}$$

$$= \arg\max_{\alpha \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha, U_i = 0, U_j = 0) \tag{24}$$

$$\overset{(a)}{=} \arg\max_{\alpha \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha, U_j = 0). \tag{25}$$

Similarly,

$$\mathcal{E}_j((i, x_i)) = \arg\max_{\alpha \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha, Y = (i, x_i)) \tag{26}$$

$$= \arg\max_{\alpha \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha, U_j = 0, U_i = 1, X_i = x_i) \tag{27}$$

$$\overset{(b)}{=} \arg\max_{\alpha \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha, U_j = 0) \tag{28}$$

$$= \mathcal{E}_j(\varnothing). \tag{29}$$

The equalities (a) and (b) follow from Assumption 1 and the fact that X_1 and X_2 are independent. Consequently, the MAP estimator $\hat{X}_j = \mathcal{E}_j(Y)$ leads to:

$$\mathbf{P}(X_j \neq \hat{X}_j | U_i = 1, U_j = 0) = \mathbf{P}(X_j \neq \hat{X}_j | U_i = 0, U_j = 0). \tag{30}$$

Finally, given that $(U_i = 0, U_j = 1)$ and $Y = (j, X_j)$ define the same event, we also have:

$$\mathbf{P}(X_j \neq \hat{X}_j | U_i = 0, U_j = 1) = 0. \tag{31}$$

Using the law of total probability, Eqs. (30) and (31), we rewrite $\rho_{\tilde{\mathcal{U}}_j}$ as follows:

$$\rho_{\tilde{\mathcal{U}}_j} = \mathbf{P}(X_j \neq \hat{X}_j, U_j = 1 | U_i = 1) \tag{32}$$

$$= \mathbf{P}(X_j \neq \mathcal{E}_j(\mathfrak{C}), U_j = 1). \tag{33}$$

Using the definition of MAP estimator and expressing the result in terms of the policy $\tilde{\mathcal{U}}_j$, we have

$$\rho_{\tilde{\mathcal{U}}_j} = \mathbf{P}(U_j = 1) - \max_{\alpha_j \in \mathbb{X}_j} \ \tilde{\mathcal{U}}_j(\alpha_j) p_{X_j}(\alpha_j). \tag{34}$$

Following similar steps, we can show that the off-set term $\theta_{\tilde{\mathcal{U}}_j}$ is given by:

$$\theta_{\tilde{\mathcal{U}}_j} = \mathbf{P}(X_j \neq \mathcal{E}_j(\varnothing), U_j = 0), \tag{35}$$

which expressed in terms of $\tilde{\mathcal{U}}_j$ is

$$\theta_{\tilde{\mathcal{U}}_j} = \mathbf{P}(U_j = 0) - \max_{x \in \mathbb{X}_j} (1 - \tilde{\mathcal{U}}_j(x)) p_{X_j}(x). \tag{36}$$

The proof is concluded by noticing that Eqs. (34) and (36) imply Eqs. (20) and (21), respectively. ■

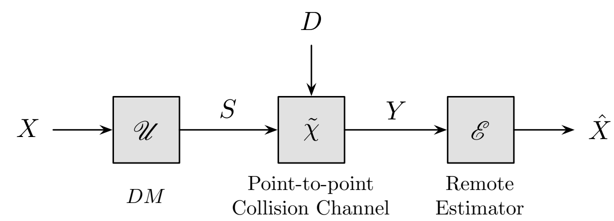

Fig. 2. An equivalent single DM estimation problem over a collision channel.

B. An equivalent single DM subproblem

In order to use the person-by-person approach to proving Theorem 1, we need to consider the subproblem of optimizing the transmission policy of one sensor, which we call DM, while assuming that the policy of $\widetilde{DM}$, representing the other sensor, is given and fixed. From the perspective of DM, the problem is depicted in Fig. 2. Here, DM observes a random variable X and must decide whether to attempt a transmission. A Bernoulli random variable D, which is independent of X, accounts for the effect that transmission attempts by $\widetilde{DM}$ have on the occurrence of collisions. The contribution of the policy of $\widetilde{DM}$ towards the cost is quantified in Eq. (17). Before we state the subproblem precisely in Problem 3, we proceed with a few definitions.

In order to emphasize the fact that D, which is determined by the fixed policy for $\widetilde{DM}$, is now a given Bernoulli random variable that can be viewed by DM as a source of randomization inherent to the channel, we adopt the following definition:

Definition 9 (Stochastic point-to-point collision channel): Let D be a given Bernoulli random variable with parameter β for which $\mathbf{P}(D = 1) = \beta$. An associated point-to-point collision channel with input S and output $Y = \tilde{\chi}(S, D)$ is specified by the following map:

$$\tilde{\chi}(s, d) \overset{\text{def}}{=} \begin{cases} \varnothing & \text{if } s = \varnothing \\ s & \text{if } s \neq \varnothing,\ d = 0 \\ \mathfrak{C} & \text{if } s \neq \varnothing,\ d = 1. \end{cases} \tag{37}$$

where s is in the input alphabet $\mathbb{S} \overset{\text{def}}{=} \mathbb{X} \cup \{\varnothing\}$. The output alphabet of the channel is $\mathbb{Y} \overset{\text{def}}{=} \mathbb{X} \cup \{\mathfrak{C}\}$.

The input to the channel is governed by DM according to:

$$s \overset{\text{def}}{=} \begin{cases} x & \text{if } u = 1 \\ \varnothing & \text{if } u = 0, \end{cases} \tag{38}$$

The following is the probability of attempting a transmission for a given measurement α:

$$\mathbf{P}(U = 1 | X = \alpha) = \mathscr{U}(\alpha), \quad \mathscr{U} \in \mathbb{U}, \alpha \in \mathbb{X} \tag{39}$$

where $\mathscr{U}$ is the transmission policy used by the DM.

Assumption 2: We assume that the randomization that generates U according to Eq. (39), is such that D is independent from the pair (U, X).

Finally, based on Eq. (17), the cost to be minimized by DM and the remote estimator is as follows:

$$\mathscr{J}_A(\mathscr{U}) \overset{\text{def}}{=} \mathbf{P}(X \neq \hat{X}) + \varrho \mathbf{P}(U = 1). \tag{40}$$

where $\mathscr{U}$ represents a transmission policy for DM and ϱ can be viewed as a communication cost induced by $\widetilde{DM}$.

Problem 3: Consider that β in $[0,1]$, real non-negative ϱ and a discrete random variable X are given. Find a policy $\mathscr{U} \in \mathbb{U}$ that minimizes the cost $\mathscr{J}_A(\mathcal{U})$ in Eq. (40), subject to Eq. (37) with $\mathbf{P}(D=1)=\beta$ and that the estimate $\hat{X}=\mathscr{E}(Y)$ is generated according to the following MAP rule:

$$\mathscr{E}(y) \overset{\text{def}}{=} \arg\max_{\alpha\in\mathbb{X}} \mathbf{P}(X=\alpha|Y=y), \quad y\in\mathbb{Y} \tag{41}$$

We will provide a solution to Problem 3 using the following two lemmata.

Lemma 1: The cost $\mathscr{J}_A(\mathscr{U})$ is concave on $\mathbb{U}$.

Proof: Using the law of total probability, we rewrite the cost $\mathscr{J}_A(\mathscr{U})$ as:

$$\begin{aligned}\mathscr{J}_A(\mathscr{U}) = \Big(\beta\mathbf{P}(X\neq\hat{X}|U=1,D=1)+\varrho\Big)\,\mathbf{P}(U=1)\\ +\mathbf{P}(X\neq\hat{X}|U=0)\mathbf{P}(U=0).\end{aligned} \tag{42}$$

Simplifying this expression using the relationships developed in the previous sections, we get:

$$\begin{aligned}\mathscr{J}_A(\mathscr{U}) =\ & 1+(\varrho+\beta-1)\mathbf{P}(U=1)\\ &-\mathbf{P}(X=\mathscr{E}(\varnothing)|U=0)\mathbf{P}(U=0)\\ &-\beta\mathbf{P}(X=\mathscr{E}(\mathfrak{C})|U=1)\mathbf{P}(U=1).\end{aligned} \tag{43}$$

Using the definition of the MAP estimator, we can write the following probabilities in terms of $\mathscr{U}$:

$$\mathbf{P}(X=\mathcal{E}(\varnothing)|U=0)=\max_{\alpha\in\mathbb{X}}\frac{(1-\mathscr{U}(\alpha))p(\alpha)}{\mathbf{P}(U=0)} \tag{44}$$

and

$$\mathbf{P}(X=\mathcal{E}(\mathfrak{C})|U=1)=\max_{\alpha\in\mathbb{X}}\frac{\mathscr{U}(\alpha)p(\alpha)}{\mathbf{P}(U=1)}. \tag{45}$$

Finally, the cost can be rewritten as follows:

$$\begin{aligned}\mathscr{J}_A(\mathscr{U}) =\ & 1+(\varrho+\beta-1)\sum_{\alpha\in\mathbb{X}}\mathscr{U}(\alpha)p(\alpha)\\ &-\max_{\alpha\in\mathbb{X}}(1-\mathscr{U}(\alpha))p(\alpha)-\beta\max_{\alpha\in\mathbb{X}}\mathscr{U}(\alpha)p(\alpha).\end{aligned} \tag{46}$$

The proof is concluded by using standard arguments found in [24, Ch. 3] to establish the concavity of Eq. (46). ■

Lemma 2: For $\beta\in[0,1]$ and $\varrho\geq 0$, the following policy minimizes $\mathscr{J}_A(\mathscr{U})$:

$$\mathscr{U}^\star_{\beta,\varrho} = \begin{cases}\mathcal{V}^{\bar{1}}_X & \text{if} \quad 0\leq\varrho\leq 1-\beta\\ \mathcal{V}^{\bar{2}}_X & \text{if} \quad 1-\beta<\varrho\leq 1\\ \mathcal{V}^{\varnothing}_X & \text{otherwise.}\end{cases} \tag{47}$$

Proof: Since $\mathscr{J}_A(\mathscr{U})$ is continuous and $\mathbb{U}$ is compact with respect to the weak* topology[2], a minimizer exists [25]. Due to the concavity of $\mathscr{J}_A(\mathscr{U})$ established in Lemma 1, the minimizer must lie on the boundary of the feasible set. Moreover, the search can be further constrained to the corners of the $|\mathbb{X}|$-dimensional hypercube that describes the feasible set and this implies that Problem 3 admits an optimal deterministic policy. Hence, it suffices to optimize with respect to policies $\mathscr{U}$ that take values in $\{0,1\}$. For each such policy, we use the alphabet partitions $\mathbb{X}=\mathbb{X}^{\mathscr{U},0}\cup\mathbb{X}^{\mathscr{U},1}$ defined as follows:

$$\mathbb{X}^{\mathscr{U},k}\overset{\text{def}}{=}\{\alpha\in\mathbb{X}\mid\mathscr{U}(\alpha)=k\},\ k\in\{0,1\}. \tag{48}$$

We proceed to finding an optimal deterministic policy by solving the equivalent problem of searching for an optimal partition. In spite of the fact that the number of partitions grows exponentially with $|\mathbb{X}|$, as we show next, we can use the cost structure to render the search for an optimal partition tractable. We start by using the partitions to rewrite the cost as follows:

$$\begin{aligned}\mathscr{J}_A(\mathscr{U}) =\ & 1+(\varrho+\beta-1)\sum_{\alpha\in\mathbb{X}^{\mathscr{U},1}}p_X(\alpha)\\ &-\max_{\alpha\in\mathbb{X}^{\mathscr{U},0}}p_X(\alpha)-\beta\max_{\alpha\in\mathbb{X}^{\mathscr{U},1}}p_X(\alpha).\end{aligned} \tag{49}$$

We will obtain a lower bound that holds for every deterministic policy $\mathscr{U}\in\mathbb{U}$, and show that $\mathscr{U}^\star_{\beta,\varrho}$ always achieves it.

First, consider the case when $\varrho>1-\beta$. Using the inequality below

$$\sum_{\alpha\in\mathbb{X}^{\mathscr{U},1}}p_X(\alpha)\geq\max_{\alpha\in\mathbb{X}^{\mathscr{U},1}}p_X(\alpha) \tag{50}$$

we conclude that the cost satisfies the following lower bound:

$$\begin{aligned}\mathscr{J}_A(\mathscr{U})\geq 1-(1-\varrho)\max_{\alpha\in\mathbb{X}^{\mathscr{U},1}}p_X(\alpha)\\ -\max_{\alpha\in\mathbb{X}^{\mathscr{U},0}}p_X(\alpha).\end{aligned} \tag{51}$$

The right hand side of the inequality above can be minimized by assigning $x_{[1]}$ to the set $\mathbb{X}^{\mathscr{U},0}$. If $1-\varrho\geq 0$, we assign $x_{[2]}$ to the set $\mathbb{X}^{\mathscr{U},1}$, otherwise we set $\mathbb{X}^{\mathscr{U},1}=\emptyset$. Therefore, we obtain the following lower bound for the cost:

$$\mathscr{J}_A(\mathscr{U})\geq 1-\max\{0,1-\varrho\}q_X(2)-q_X(1). \tag{52}$$

When $1-\beta\leq\varrho$, this lower bound is met with equality by the policy $\mathscr{U}^\star_{\beta,\varrho}$, for which the cost is given by:

$$\mathscr{J}_A(\mathscr{U}^\star_{\beta,\varrho})=\begin{cases}1-q_X(1) & \text{if } \varrho>1\\ 1-(1-\varrho)q_X(2)-q_X(1) & \text{otherwise.}\end{cases} \tag{53}$$

Similarly, when $0\leq\varrho\leq 1-\beta$, we have

$$\sum_{\alpha\in\mathbb{X}^{\mathscr{U},1}}p_X(\alpha)\leq 1-\max_{\alpha\in\mathbb{X}^{\mathscr{U},0}}p_X(\alpha). \tag{54}$$

Therefore, for every $\mathscr{U}$ in $\mathbb{U}$, we establish the following lower bound on the cost:

$$\begin{aligned}\mathscr{J}_A(\mathscr{U})\geq(\varrho+\beta)(1-\max_{\alpha\in\mathbb{X}^{\mathscr{U},0}}p_X(\alpha))\\ -\beta\max_{\alpha\in\mathbb{X}^{\mathscr{U},1}}p_X(\alpha).\end{aligned} \tag{55}$$

The right hand side of the inequality above can be minimized by assigning $x_{[1]}$ to the set $\mathbb{X}^{\mathscr{U},0}$. If $1-\varrho\geq 0$, we assign $x_{[2]}$ to the set $\mathbb{X}^{\mathscr{U},1}$. Therefore, we obtain the following lower bound for the cost:

$$\mathscr{J}_A(\mathscr{U})\geq(\varrho+\beta)(1-q_X(1))-\beta q_X(2). \tag{56}$$

[2]This technical detail can be ignored when $|\mathbb{X}|<\infty$.

The policy $\mathscr{U}^{\star}_{\beta,\varrho}$ achieves this lower bound, as the following calculation shows:

$$\begin{aligned}\mathscr{J}_A(\mathscr{U}^{\star}_{\beta,\varrho}) &= 1 + (\varrho+\beta-1)\sum_{\alpha\in\mathbb{X}\setminus\{x_{[1]}\}} p_X(\alpha) \\ &\quad - q_X(1) - \beta \max_{\alpha\in\mathbb{X}\setminus\{x_{[1]}\}} p_X(\alpha) \\ &= (\varrho+\beta)(1-q_X(1)) - \beta q_X(2). \quad (57)\end{aligned}$$

■

Remark 4: Lemma 2 provides a solution to Problem 3 described only in terms of β, ϱ and the two most likely outcomes of X. As a particular case, when β is zero and ϱ is in $[0,1]$, the optimal policy is

$$\mathscr{U}^{\star}_{0,\varrho}(\alpha) = \begin{cases} 0 & \text{if } \alpha = x_{[1]} \\ 1 & \text{otherwise.} \end{cases} \quad (58)$$

This result is related to a similar problem solved by Imer and Basar in [16].

C. Globally optimal solutions and Proof of Theorem 1

We will now apply the results in Section III-B to reduce the search space of possible optimal strategies for each DM in Problem 1. The strategy is to use a person-by-person optimality approach together with Lemma 2.

Proof of Theorem 1: Consider the cost $\mathcal{J}_A(\mathcal{U}_1,\mathcal{U}_2)$ in Problem 1. Arbitrarily fixing the policy $\tilde{\mathcal{U}}_2$ of DM_2, we have:

$$\mathcal{J}_A(\mathcal{U}_1,\tilde{\mathcal{U}}_2) \propto \mathbf{P}(X_1 \neq \hat{X}_1) + \frac{\eta_2}{\eta_1}(\rho_{\tilde{\mathcal{U}}_2}\mathbf{P}(U_1=1) + \theta_{\tilde{\mathcal{U}}_2}). \quad (59)$$

The problem of minimizing $\mathcal{J}_A(\mathcal{U}_1,\tilde{\mathcal{U}}_2)$ over $\mathcal{U}_1 \in \mathbb{U}_1$ is equivalent to solving an instance of Problem 3 with parameters ϱ and β selected as follows:

$$\varrho = \frac{\eta_2}{\eta_1}\rho_{\tilde{\mathcal{U}}_2} \quad \text{and} \quad \beta = \mathbf{P}(U_2=1). \quad (60)$$

Hence, from Lemmas 1 and 2, for each policy $\tilde{\mathcal{U}}_2$ in $\mathbb{U}_2$ there is at least one choice for $\mathcal{U}_1^{\star}$ in $\{\mathcal{V}^{\bar{1}}_{X_1}, \mathcal{V}^{2}_{X_1}, \mathcal{V}^{\varnothing}_{X_1}\}$ for which $\mathcal{J}_A(\mathcal{U}_1^{\star},\tilde{\mathcal{U}}_2) \leq \mathcal{J}_A(\mathcal{U}_1,\tilde{\mathcal{U}}_2)$ holds for any $\mathcal{U}_1$ in $\mathbb{U}_1$. Since this is true regardless of our choice of $\tilde{\mathcal{U}}_2$ and it also holds if we were to fix the policy of DM_1 and optimize $\mathcal{U}_2$, we conclude that given any person-by-person optimal pair $(\mathcal{U}_1^{\star},\mathcal{U}_2^{\star})$, there is a pair $(\check{\mathcal{U}}_1,\check{\mathcal{U}}_2)$ in $\{\mathcal{V}^{\bar{1}}_{X_1}, \mathcal{V}^{2}_{X_1}, \mathcal{V}^{\varnothing}_{X_1}\} \times \{\mathcal{V}^{\bar{1}}_{X_2}, \mathcal{V}^{2}_{X_2}, \mathcal{V}^{\varnothing}_{X_2}\}$, for which $\mathcal{J}_A(\check{\mathcal{U}}_1,\check{\mathcal{U}}_2) \leq \mathcal{J}_A(\mathcal{U}_1^{\star},\mathcal{U}_2^{\star})$ holds. The proof of Theorem 1 is complete once we recall that every globally optimal solution is also person-by-person optimal. ■

Remark 5: There may be other optimal solutions that do not have the same structure of the policies in Theorem 1. Note that the performance of an optimal remote estimation system is determined by the probabilities of the most likely outcomes of X_1 and X_2. Also, the optimal performance of a system with binary observations is always zero, i.e., independent binary observations can be estimated perfectly from the output of the collision channel with two users. The pair of globally optimal solutions described in the motivating example of Section II-C also fits in the structure of the globally optimal solutions of Theorem 1.

TABLE I
VALUE OF THE COST FUNCTION $\mathcal{J}_A(\mathcal{U}_1,\mathcal{U}_2)$ AT EACH OF THE NINE CANDIDATE SOLUTIONS SPECIFIED IN THEOREM 1.

m	$(\mathcal{U}_1,\mathcal{U}_2)$	$\mathcal{J}_A^{(m)} \overset{\text{def}}{=} \mathcal{J}_A(\mathcal{U}_1,\mathcal{U}_2)$
1	$(\mathcal{V}^{\bar{1}}_{X_1},\mathcal{V}^{\bar{1}}_{X_2})$	$\eta_1 t_{X_1}(1-q_{X_2}(1)) + \eta_2 t_{X_2}(1-q_{X_1}(1))$
2	$(\mathcal{V}^{\bar{1}}_{X_1},\mathcal{V}^{2}_{X_2})$	$\eta_1 t_{X_1} q_{X_2}(2) + \eta_2 t_{X_2}$
3	$(\mathcal{V}^{2}_{X_1},\mathcal{V}^{\bar{1}}_{X_2})$	$\eta_1 t_{X_1} + \eta_2 t_{X_2} q_{X_1}(2)$
4	$(\mathcal{V}^{\bar{1}}_{X_1},\mathcal{V}^{\varnothing}_{X_2})$	$\eta_2(1-q_{X_2}(1))$
5	$(\mathcal{V}^{\varnothing}_{X_1},\mathcal{V}^{\bar{1}}_{X_2})$	$\eta_1(1-q_{X_1}(1))$
6	$(\mathcal{V}^{2}_{X_1},\mathcal{V}^{2}_{X_2})$	$\eta_1 t_{X_1} + \eta_2 t_{X_2}$
7	$(\mathcal{V}^{2}_{X_1},\mathcal{V}^{\varnothing}_{X_2})$	$\eta_1 t_{X_1} + \eta_2(1-q_{X_2}(1))$
8	$(\mathcal{V}^{\varnothing}_{X_1},\mathcal{V}^{2}_{X_2})$	$\eta_1(1-q_{X_1}(1)) + \eta_2 t_{X_2}$
9	$(\mathcal{V}^{\varnothing}_{X_1},\mathcal{V}^{\varnothing}_{X_2})$	$\eta_1(1-q_{X_1}(1)) + \eta_2(1-q_{X_2}(1))$

We proceed to evaluating the performance of each of the nine candidate policy-pairs listed in Theorem 1 using the expressions in Eqs. (18) and (19), and the following quantity

$$t_{X_i} \overset{\text{def}}{=} 1 - q_{X_i}(1) - q_{X_i}(2), \ i \in \{1,2\}, \quad (61)$$

where q_W is as defined in Eq. (13).

- If $\mathcal{U}_i = \mathcal{V}^{\bar{1}}_{X_i}$, then

$$\mathbf{P}(U_i=1) = 1 - q_{X_i}(1) \quad (62)$$

$$\rho_{\mathcal{U}_i} = t_{X_i} \quad \text{and} \quad \theta_{\mathcal{U}_i} = 0. \quad (63)$$

- If $\mathcal{U}_i = \mathcal{V}^{2}_{X_i}$, then

$$\mathbf{P}(U_i=1) = q_{X_i}(2) \quad (64)$$

$$\rho_{\mathcal{U}_i} = 0 \quad \text{and} \quad \theta_{\mathcal{U}_i} = t_{X_i}. \quad (65)$$

- If $\mathcal{U}_i = \mathcal{V}^{\varnothing}_{X_i}$, then

$$\mathbf{P}(U_i=1) = 0 \quad (66)$$

$$\rho_{\mathcal{U}_i} = 0 \quad \text{and} \quad \theta_{\mathcal{U}_i} = 1 - q_{X_i}(1). \quad (67)$$

We construct Table I, which lists the cost evaluated for all candidate policy pairs. It can be verified by inspection that the policy pairs for which m is $6, 7, 8$ and 9 are always outperformed by at least one of the others. This observation leads to the following corollary.

Corollary 1: The optimal cost obtained from solving Problem 1 is given by

$$\mathcal{J}_A^{\star} = \min_{1\leq m\leq 5} \mathcal{J}_A^{(m)} \quad (68)$$

where $\mathcal{J}_A^{m}$, for m in $\{1,\ldots,9\}$, is specified in Table I.

D. Examples

We explore the role that p_{X_1} and p_{X_2} have in determining which of the 5 policy pairs ($m = 1$ through 5 in I) is globally optimal. In the examples below, we assume[3] that $\eta_1 = \eta_2$,

[3] In this case, the weights η_1 and η_2 are irrelevant and we may assume that they are both equal to 1.

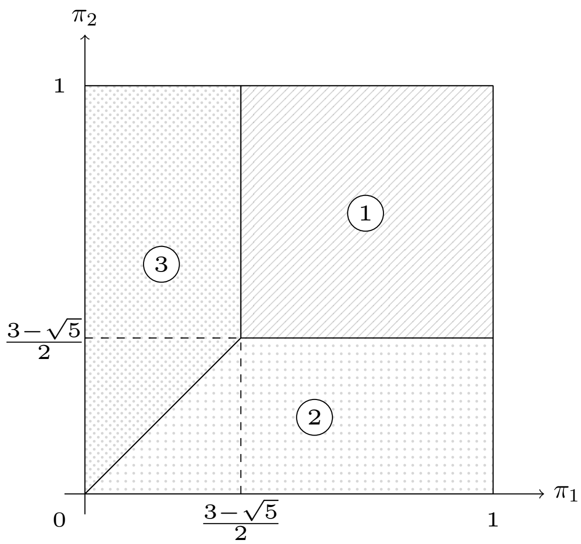

Fig. 3. Partition of the parameter space indicating where each policy pair is globally optimal for Example 2. The circled number corresponds to m in Table I.

which further reduces our search to policy pairs $m = 1, 2$ and 3. We will use the following quantities:

$$\mathcal{J}_A^{(2)} - \mathcal{J}_A^{(3)} = -t_{X_1}(1 - q_{X_2}(2)) + t_{X_2}(1 - q_{X_1}(2)) \quad (69)$$

$$\mathcal{J}_A^{(2)} - \mathcal{J}_A^{(1)} = t_{X_2}(q_{X_1}(1) - t_{X_1}) \quad (70)$$

$$\mathcal{J}_A^{(3)} - \mathcal{J}_A^{(1)} = t_{X_1}(q_{X_2}(1) - t_{X_2}). \quad (71)$$

Example 1 (Uniformly distributed observations): For uniformly distributed observations, we have

$$p_{X_i}(x) = \frac{1}{N_i}, \ x = 1, 2, \cdots, N_i. \quad (72)$$

Hence, the probabilities of the two most likely outcomes are

$$q_{X_i}(1) = q_{X_i}(2) = \frac{1}{N_i} \quad (73)$$

and the aggregate probability of all other outcomes is given by

$$t_{X_i} = 1 - \frac{2}{N_i}, \ i \in \{1, 2\}. \quad (74)$$

Without loss of generality, we assume that $N_1, N_2 \geq 3$ and $N_1 \leq N_2$. Since

$$\mathcal{J}_A^{(2)} - \mathcal{J}_A^{(3)} = \frac{1}{N_1} - \frac{1}{N_2} \geq 0 \quad (75)$$

$$\mathcal{J}_A^{(3)} - \mathcal{J}_A^{(1)} = \left(1 - \frac{2}{N_1}\right) \times \left(\frac{3}{N_2} - 1\right) \leq 0, \quad (76)$$

our assumptions imply that $\mathcal{J}_A^\star = \mathcal{J}_A^{(3)}$ and the pair of policies corresponding to $m = 3$ is globally optimal.

Example 2 (Geometrically distributed observations):

For geometrically distributed observations with parameters π_1 and π_2, we have

$$p_{X_i}(x) = (1 - \pi_i)^x \pi_i, \ x \geq 0, \ i \in \{1, 2\} \quad (77)$$

The probabilities of the two most likely outcomes for each sensor are:

$$q_{X_i}(1) = \pi_i, \ i \in \{1, 2\} \quad (78)$$

$$q_{X_i}(2) = (1 - \pi_i)\pi_i, \ i \in \{1, 2\} \quad (79)$$

and the aggregate probability of all other outcomes is:

$$t_{X_i} = (1 - \pi_i)^2, \ i \in \{1, 2\}. \quad (80)$$

Note that $\mathcal{J}_A^{(1)}$ is less than or equal to $\mathcal{J}_A^{(2)}$ and $\mathcal{J}_A^{(3)}$ if and only if $q_{X_i}(1) \geq t_{X_i}$, for $i \in \{1, 2\}$, or equivalently, if the following holds:

$$-\pi_i^2 + 3\pi_i - 1 \geq 0, \ i \in \{1, 2\}. \quad (81)$$

Also, $\mathcal{J}_A^{(2)}$ is less than or equal to $\mathcal{J}_A^{(3)}$ if and only if the following holds:

$$(1 - \pi_2)^2 \pi_1 \leq (1 - \pi_1)^2 \pi_2, \quad (82)$$

which is satisfied if $\pi_1 \leq \pi_2$. This yields the partitioning of the parameter space $(\pi_1, \pi_2) \in [0, 1]^2$ into the three regions depicted in Fig. 3. Each region is labelled according to the policy pair that is optimal within it.

Example 3 (Poisson distributed observations): For Poisson distributed observations with parameters λ_1 and λ_2, which we assume here to be both greater than 1, the probabilities of the two most likely outcomes are

$$q_{X_i}(1) = q_{X_i}(2) = \frac{\lambda_i^{\lfloor \lambda_i \rfloor}}{\lfloor \lambda_i \rfloor !} e^{-\lambda_i} \quad (83)$$

and the aggregate probability of all other outcomes is:

$$t_{X_i} = 1 - 2\frac{\lambda_i^{\lfloor \lambda_i \rfloor}}{\lfloor \lambda_i \rfloor !} e^{-\lambda_i}, \ i \in \{1, 2\}. \quad (84)$$

Using the same argument as in the previous example, we note that $\mathcal{J}_A^{(1)}$ is less than or equal to $\mathcal{J}_A^{(2)}$ and $\mathcal{J}_A^{(3)}$ if and only if $q_{X_i}(1) \geq t_{X_i}$ is satisfied, or equivalently, the following holds:

$$\frac{\lambda_i^{\lfloor \lambda_i \rfloor}}{\lfloor \lambda_i \rfloor !} e^{-\lambda_i} \geq \frac{1}{3}, \ i \in \{1, 2\}. \quad (85)$$

In order to check whether Eg. (85) holds, we define the following function

$$\mathcal{F}(\lambda) \stackrel{\text{def}}{=} \frac{\lambda^{\lfloor \lambda \rfloor}}{\lfloor \lambda \rfloor !} e^{-\lambda} - \frac{1}{3}, \quad (86)$$

It can be shown that $\mathcal{F}(\lambda)$ is greater than or equal to zero if and only if $0 < \lambda \leq \bar{\lambda} \approx 1.5121$. Hence, we conclude that Eq. (85) holds if and only if $1 \leq \lambda_i \leq \bar{\lambda}$ for $i \in \{1, 2\}$. Finally, $\mathcal{J}_A^{(2)} \leq \mathcal{J}_A^{(3)}$ holds if and only if

$$\frac{\lambda_2^{\lfloor \lambda_2 \rfloor}}{\lfloor \lambda_2 \rfloor !} e^{-\lambda_2} \geq \frac{\lambda_1^{\lfloor \lambda_1 \rfloor}}{\lfloor \lambda_1 \rfloor !} e^{-\lambda_1}, \quad (87)$$

which is satisfied when $\lambda_1 \geq \lambda_2$.

Example 4 (Identically distributed observations):

When the observations are identically distributed, i.e., $p_{X_1} = p_{X_2}$, and $\mathbb{X}_1 = \mathbb{X}_2$, we have:

$$\mathcal{J}_A^{(2)} = \mathcal{J}_A^{(3)} = (1 - q_{X_1}(1) - q_{X_1}(2))(1 + q_{X_1}(2)) \quad (88)$$

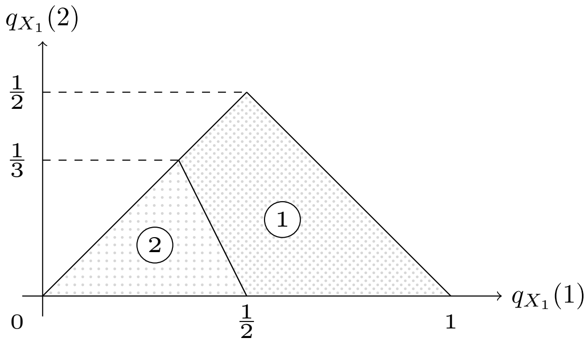

Fig. 4. Partition of the parameter space indicating where each policy pair is globally optimal for Example 4. The circled number corresponds to m in Table I.

and

$$\mathcal{J}_A^{(1)} = (1 - q_{X_1}(1) - q_{X_1}(2))(2 - 2q_{X_1}(1)) \tag{89}$$

Therefore, $\mathcal{J}_A^{(2)} \leq \mathcal{J}_A^{(1)}$ if and only if

$$2q_{X_1}(1) + q_{X_1}(2) \leq 1. \tag{90}$$

Recalling that $q_{X_1}(1) \geq q_{X_1}(2)$, we have the partitioning of the parameter space $[0,1]^2$ according to Fig. 4.

IV. Solution to Problem 2

In this section, we provide a proof for Theorem 2, which characterizes transmission policies for the sensors that minimize the probability that either $\hat{X}_1$ or $\hat{X}_2$, or both, give an incorrect estimate. The cost, which was initially defined as Eq. (7), is now rewritten as follows:

$$\mathcal{J}_B(\mathcal{U}_1, \mathcal{U}_2) = \mathbf{P}(W \neq \hat{W}). \tag{91}$$

where $W \overset{\text{def}}{=} (X_1, X_2)$ and the estimate $\hat{W} \overset{\text{def}}{=} \mathcal{E}(Y)$ is determined by the following MAP rule:

$$\mathcal{E}(y) = \arg\max_{\omega \in \mathbb{W}} \mathbf{P}(W = \omega | Y = y), \ y \in \mathbb{Y} \tag{92}$$

where $\mathbb{W} = \mathbb{X}_1 \times \mathbb{X}_2$.

The overall proof strategy is centered on the characterization of globally optimal strategies via the person-by-person optimality approach. This is possible in spite of the fact that the cost $\mathcal{J}_B(\mathcal{U}_1, \mathcal{U}_2)$ does not admit the additive decomposition used in Problem 1.

We start by stating two propositions that provide identities useful in Section IV-A, where we use the total probability law to rewrite the cost in a convenient way.

Proposition 2: The following holds for Problem 2:

$$\begin{aligned}\mathbf{P}(W = \hat{W} | Y = \mathfrak{C}) = & \max_{\alpha_1 \in \mathbb{X}_1} \mathbf{P}(X_1 = \alpha_1 | U_1 = 1) \\ & \times \max_{\alpha_2 \in \mathbb{X}_2} \mathbf{P}(X_2 = \alpha_2 | U_2 = 1);\end{aligned} \tag{93}$$

and

$$\begin{aligned}\mathbf{P}(W = \hat{W} | Y = \varnothing) = & \max_{\alpha_1 \in \mathbb{X}_1} \mathbf{P}(X_1 = \alpha_1 | U_1 = 0) \\ & \times \max_{\alpha_2 \in \mathbb{X}_2} \mathbf{P}(X_2 = \alpha_2 | U_2 = 0).\end{aligned} \tag{94}$$

Proof: The conditional probability of a correct estimate conditioned on the event of a collision can be computed as:

$$\begin{aligned}&\mathbf{P}(W = \hat{W} | Y = \mathfrak{C}) \\ &\quad \overset{(a)}{=} \max_{\omega \in \mathbb{W}} \mathbf{P}(W = \omega | Y = \mathfrak{C}) && (95)\\ &\quad \overset{(b)}{=} \max_{\omega \in \mathbb{W}} \mathbf{P}(W = \omega | U_1 = 1, U_2 = 1) && (96)\\ &\quad \overset{(c)}{=} \max_{\alpha_1 \in \mathbb{X}_1} \mathbf{P}(X_1 = \alpha_1 | U_1 = 1) \\ &\qquad \times \max_{\alpha_2 \in \mathbb{X}_2} \mathbf{P}(X_2 = \alpha_2 | U_2 = 1). && (97)\end{aligned}$$

where (a) follows from our definition of $\hat{W}$; the equality in (b) results from the fact that the events $(Y = \mathfrak{C})$ and $(U_1 = 1, U_2 = 1)$ are identical; finally, (c) follows from the fact that (U_1, X_1) and (U_2, X_2) are independent.

The proof of the second equality can be derived from the fact that the events $(Y = \varnothing)$ and $(U_1 = 0, U_2 = 0)$ are identical, followed by the same steps as before. ■

Proposition 3: The following holds for Problem 2:

$$\mathbf{P}(W = \hat{W} | Y = (i, X_i)) = \max_{\alpha_j \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha_j | U_j = 0) \tag{98}$$

with $i, j \in \{1, 2\}$ and $i \neq j$.

Proof: It suffices to show that for every $\alpha_i \in \mathbb{X}_i$, the following equalities hold:

$$\begin{aligned}&\mathbf{P}(W = \hat{W} | Y = (i, \alpha_i)) \\ &\quad \overset{(a)}{=} \max_{\omega \in \mathbb{W}} \mathbf{P}(W = \omega | Y = (i, \alpha_i)) && (99)\\ &\quad \overset{(b)}{=} \max_{\omega \in \mathbb{W}} \mathbf{P}(W = \omega | U_i = 1, X_i = \alpha_i, U_j = 0) && (100)\\ &\quad \overset{(c)}{=} \max_{\alpha_j \in \mathbb{X}_j} \mathbf{P}(X_j = \alpha_j | U_j = 0). && (101)\end{aligned}$$

Where (a) follows from our definition of $\hat{W}$; the equality in (b) follows from the fact that the events $(Y = (i, \alpha_i))$ and $(U_i = 1, X_i = \alpha_i, U_j = 0)$ are equivalent; and (c) follows from the fact that (U_1, X_1) and (U_2, X_2) are independent. ■

A. An equivalent single DM subproblem

Here, we adopt an approach analogous to the one used in Section III-B to prove Theorem 1. In particular, we proceed to providing preliminary results that we will be used in Section IV-B to prove Theorem 2 via the person-by-person approach.

A key step is to characterize Problem 2 from the viewpoint of one agent, when the transmission policy of the other is given and fixed. Unlike Problem 1, in which the cost structure from the viewpoint of each sensor allowed us to make an analogy with a remote estimation problem subject to communication costs, Problem 2 does not admit an insightful decomposition. Fortunately, we can still use the same techniques applied to a less convenient cost.

Similar to the approach adopted in Section III-B, for a given discrete random variable X, we define $\mathbb{U}$ to be the class of all functions $\mathscr{U}$ with domain $\mathbb{X}$ taking values in $[0,1]$. Elements of $\mathbb{U}$ represent policies that govern transmission in the same manner described in Section III-B, with the difference that we now consider the cost $\mathscr{J}_B : \mathbb{U} \to \mathbb{R}_{\geq 0}$ defined as follows:

$$\mathscr{J}_B(\mathscr{U}) \stackrel{\text{def}}{=} 1 - \tau \max_{\alpha \in \mathbb{X}} \mathscr{U}(\alpha) p_X(\alpha) - \varrho \sum_{\alpha \in \mathbb{X}} \mathscr{U}(\alpha) p_X(\alpha) - (\varrho + \beta) \max_{\alpha \in \mathbb{X}} (1 - \mathscr{U}(\alpha)) \, p_X(\alpha). \quad (102)$$

where ϱ, τ and β are non-negative constants.

Lemma 3: The cost $\mathscr{J}_B$ is concave on $\mathbb{U}$.

Proof: The proof follows from standard arguments that can be found in [24, Ch. 3]. ■

Lemma 4: Let X be a given discrete random variable. If $\tau \leq \beta$ then $\mathcal{V}_X^{\bar{1}}$ minimizes $\mathscr{J}_B$.

Proof: From Lemma 3, the cost is concave in $\mathscr{U}$. Therefore, without loss in optimality, we can constrain the optimization to the class of deterministic strategies. For any deterministic policy $\mathscr{U} \in \mathbb{U}$, define

$$\mathbb{X}^{\mathscr{U},k} \stackrel{\text{def}}{=} \{\alpha \in \mathbb{X} \mid \mathscr{U}(\alpha) = k\}, \quad k \in \{0,1\}. \quad (103)$$

Constraining the policies to be deterministic and using the notation defined above, the cost becomes

$$\mathscr{J}_B(\mathscr{U}) = 1 - \tau \max_{\alpha \in \mathbb{X}^{\mathscr{U},1}} p_X(\alpha) - \varrho \sum_{\alpha \in \mathbb{X}^{\mathscr{U},1}} p_X(\alpha) - (\varrho + \beta) \max_{\alpha \in \mathbb{X}^{\mathscr{U},0}} p_X(\alpha). \quad (104)$$

Since

$$\sum_{\alpha \in \mathbb{X}^{\mathscr{U},1}} p_X(\alpha) \leq 1 - \max_{\alpha \in \mathbb{X}^{\mathscr{U},0}} p_X(\alpha), \quad (105)$$

we obtain the following inequality, which holds for every deterministic policy $\mathscr{U} \in \mathbb{U}$:

$$\mathscr{J}_B(\mathscr{U}) \geq 1 - \varrho - \tau \max_{\alpha \in \mathbb{X}^{\mathscr{U},1}} p_X(\alpha) - \beta \max_{\alpha \in \mathbb{X}^{\mathscr{U},0}} p_X(\alpha). \quad (106)$$

If $\tau \leq \beta$ then the lower bound on the right hand side of the inequality above can be minimized by assigning the symbol $x_{[1]}$ to $\mathbb{X}^{\mathscr{U},0}$ and $x_{[2]}$ to $\mathbb{X}^{\mathscr{U},1}$, yielding:

$$\mathscr{J}_B(\mathscr{U}) \geq 1 - \varrho - \tau q_X(2) - \beta q_X(1). \quad (107)$$

The lower bound in Eq. (107) is achieved when $\mathscr{U} = \mathcal{V}_X^{\bar{1}}$. ■

B. Globally optimal solution and Proof of Theorem 2

We are now equipped to present a proof for Theorem 2.

Proof of Theorem 2: In order to use the person-by-person optimality approach, we will rewrite the cost from the perspective of a single decision maker. The law of total probability and the results in Propositions 2 and 3 allow us to re-express the cost as follows:

$$\begin{aligned}
\mathcal{J}_B(\mathcal{U}_1, \mathcal{U}_2) = {} & \\
& 1 - \max_{\alpha_1 \in \mathbb{X}_1} \mathcal{U}_1(\alpha_1) p_{X_1}(\alpha_1) \max_{\alpha_2 \in \mathbb{X}_2} \mathcal{U}_2(\alpha_2) p_{X_2}(\alpha_2) \\
& - \max_{\alpha_1 \in \mathbb{X}_1} (1 - \mathcal{U}_1(\alpha_1)) p_{X_1}(\alpha_1) \max_{\alpha_2 \in \mathbb{X}_2} (1 - \mathcal{U}_2(\alpha_2)) p_{X_2}(\alpha_2) \\
& - \sum_{\alpha_1 \in \mathbb{X}_1} \mathcal{U}_1(\alpha_1) p_{X_1}(\alpha_1) \max_{\alpha_2 \in \mathbb{X}_2} (1 - \mathcal{U}_2(\alpha_2)) p_{X_2}(\alpha_2) \\
& - \sum_{\alpha_2 \in \mathbb{X}_2} \mathcal{U}_2(\alpha_2) p_{X_2}(\alpha_2) \max_{\alpha_1 \in \mathbb{X}_1} (1 - \mathcal{U}_1(\alpha_1)) p_{X_1}(\alpha_1). \quad (108)
\end{aligned}$$

We start by fixing the transmission policy of sensor 2 to an arbitrary choice $\tilde{\mathcal{U}}_2$. We use Eq. (108) to write the cost from the perspective of DM_1 as follows:

$$\begin{aligned}
\mathcal{J}_B(\mathcal{U}_1, \tilde{\mathcal{U}}_2) = {} & 1 - \tilde{\tau}_2 \max_{\alpha_1 \in \mathbb{X}_1} \mathcal{U}_1(\alpha_1) p_{X_1}(\alpha_1) \\
& - \tilde{\varrho}_2 \sum_{\alpha_1 \in \mathbb{X}_1} \mathcal{U}_1(\alpha_1) p_{X_1}(\alpha_1) \\
& - (\tilde{\varrho}_2 + \tilde{\beta}_2) \max_{\alpha_1 \in \mathbb{X}_1} (1 - \mathcal{U}_1(\alpha_1)) p_{X_1}(\alpha_1), \quad (109)
\end{aligned}$$

where

$$\tilde{\beta}_2 \stackrel{\text{def}}{=} \sum_{\alpha_2 \in \mathbb{X}_2} \tilde{\mathcal{U}}_2(\alpha_2) p_{X_2}(\alpha_2), \quad (110)$$

$$\tilde{\varrho}_2 \stackrel{\text{def}}{=} \max_{\alpha_2 \in \mathbb{X}_2} (1 - \tilde{\mathcal{U}}_2(\alpha_2)) p_{X_2}(\alpha_2) \quad (111)$$

and

$$\tilde{\tau}_2 \stackrel{\text{def}}{=} \max_{\alpha_2 \in \mathbb{X}_2} \tilde{\mathcal{U}}_2(\alpha_j) p_{X_2}(\alpha_2). \quad (112)$$

Note that for any given $\tilde{\mathcal{U}}_2$ in $\mathbb{U}_2$, we have $\tilde{\beta}_2 \geq \tilde{\tau}_2$. Hence, from Lemma 4, $\mathcal{J}_B(\mathcal{V}_{X_1}^{\bar{1}}, \tilde{\mathcal{U}}_2) \leq \mathcal{J}_B(\mathcal{U}_1, \tilde{\mathcal{U}}_2)$ holds for any $\mathcal{U}_1$ in $\mathbb{U}_1$. Given the facts that the choice for $\tilde{\mathcal{U}}_2$ was arbitrary and that we could alternatively have chosen to fix the policy of sensor DM_1 to an arbitrary selection $\tilde{\mathcal{U}}_1$ and optimized with respect to $\mathcal{U}_2$, we conclude that the policy pair $(\mathcal{V}_{X_1}^{\bar{1}}, \mathcal{V}_{X_2}^{\bar{1}})$ is globally optimal for Problem 2. ■

V. Extensions to more than two sensors

Although, thus far, our analysis has been restricted to two decision makers, the techniques and results presented can be easily extended to the case in which there are three or more sensors.

We now proceed to extending our results to allow for a team of N (possibly greater than two) sensors that access independent observations $\{X_1, \ldots, X_N\}$ and communicate over a collision channel that can only support one transmission. Let $U = (U_1, \cdots, U_N)$ denote the N-tuple of transmission decision variables, and U_{-i} denote the $N-1$ tuple obtained by excluding U_i from U. Similarly, we use $\mathcal{U} = (\mathcal{U}_1, \cdots, \mathcal{U}_N)$ to represent the N-tuple of transmission policies and $\mathcal{U}_{-i}$ is obtained by omitting $\mathcal{U}_i$ from $\mathcal{U}$.

Assumption 3: The transmission decision U_i of each DM_i is generated as a function of $\mathcal{U}_i$ in the same manner as described in Definition 1, with the evident modification that we now consider that i is in $\{1, \ldots, N\}$. We also assume that the underlying randomization that generates U from $\mathcal{U}$ is such that the pairs $\{(U_i, X_i)\}_{i=1}^N$ are mutually independent.

Assumption 4: The random elements S_1 through S_N are generated in the same manner as described in Definition 2, with the evident modification that we now consider that i is in $\{1, \ldots, N\}$.

The collision channel operates as follows:

Definition 10: The collision channel accepts inputs S_1 through S_N. The output of the collision channel is specified by the following map:

$$y = \aleph(s_1, \ldots, s_N) \stackrel{\text{def}}{=} \begin{cases} s_i & \text{if } s_i \neq \varnothing \text{ and } s_j = \varnothing,\ j \neq i \\ \varnothing & \text{if } s_i = \varnothing,\ 1 \leq i \leq N \\ (\mathfrak{C}, u) & \text{otherwise.} \end{cases} \quad (113)$$

Remark 6: Notice that $\aleph$ defined above is a natural extension of χ. The only difference is that, when there is a collision, $\aleph$ also conveys the decision vector U. Notice that both channels are equivalent when N is two because, in that case, a collision can only occur when U_1 and U_2 are both one. Hence, if the estimator is informed that a collision occurred and N is two, the additional information on U provided by $\aleph$ becomes redundant. Notice that the recovery of U at the receiver when there are collisions, as is assumed for $\aleph$, has been demonstrated empirically in [26] for N possibly greater than two.

A. *Extension of Theorem 1 to three or more sensors*

Using the same person-by-person approach of the previous sections, the result stated in Theorem 1 can be extended to the case when there are more than two sensors. In this section, we provide a sketch of the proof. We proceed to considering a version of Problem 1 in which there are N sensors, and the cost is as follows:

$$\mathcal{J}_A(\mathcal{U}) \stackrel{\text{def}}{=} \sum_{k=1}^{N} \eta_k \mathbf{P}(X_k \neq \hat{X}_k) \tag{114}$$

where η_k are positive constants.

The following is the extended version of 1 for the case when there are N sensors.

Theorem 3: There exists a globally optimal solution $(\check{\mathcal{U}}_1, \dots, \check{\mathcal{U}}_N)$ for which $\check{\mathcal{U}}_i$ is in $\{\mathcal{V}^{\bar{1}}_{X_i}, \mathcal{V}^{2}_{X_i}, \mathcal{V}^{\varnothing}_{X_i}\}$, for all i in $\{1, \dots, N\}$.

Proof: Select arbitrary i in $\{1, \dots, N\}$ and an $N-1$ tuple $\tilde{\mathcal{U}}_{-i}$, and write the cost from the perspective of DM_i as follows:

$$\mathcal{J}_A(\mathcal{U}_i, \tilde{\mathcal{U}}_{-i}) = \eta_i \mathbf{P}(X_i \neq \hat{X}_i) + \rho_{\tilde{\mathcal{U}}_{-i}} \mathbf{P}(U_i = 1) + \theta_{\tilde{\mathcal{U}}_{-i}}, \tag{115}$$

where the "communication cost" and offset terms are given by

$$\rho_{\tilde{\mathcal{U}}_{-i}} \stackrel{\text{def}}{=} \sum_{j \neq i} \eta_j \big(\mathbf{P}(X_j \neq \hat{X}_j | U_i = 1) - \mathbf{P}(X_j \neq \hat{X}_j | U_i = 0)\big) \tag{116}$$

and

$$\theta_{\tilde{\mathcal{U}}_{-i}} \stackrel{\text{def}}{=} \sum_{j \neq i} \eta_j \mathbf{P}(X_j \neq \hat{X}_j | U_i = 0). \tag{117}$$

We can conclude by inspection that minimizing the cost above with respect to DM_i is equivalent to 3 with ϱ and β given by:

$$\varrho \stackrel{\text{def}}{=} \frac{\rho_{\tilde{\mathcal{U}}_{-i}}}{\eta_i}, \tag{118}$$

$$\beta \stackrel{\text{def}}{=} 1 - \prod_{j \neq i} \mathbf{P}(U_j = 0). \tag{119}$$

Here, β is the probability that a transmission by DM_i will collide with a transmission by at least one of the other $N-1$ sensors. At this point, the proof follows from 2. ■

B. *Extension of Theorem 2 to three or more sensors*

Here, we seek to find $\mathcal{U}$ that minimizes the following cost:

$$\mathcal{J}_B(\mathcal{U}) = \mathbf{P}(W \neq \hat{W}), \tag{120}$$

where $\hat{W} = (\hat{X}_1, \dots, \hat{X}_N)$ is a N-tuple MAP estimate for $(X_1, \dots, X_N)$ obtained as follows:

$$\hat{W} = \mathcal{E}(Y) \tag{121}$$

and

$$\mathcal{E}(y) = \arg\max_{\alpha \in \mathbb{W}} \mathbf{P}((X_1, \dots, X_N) = \alpha | Y = y) \tag{122}$$

where $\mathbb{W} \stackrel{\text{def}}{=} \mathbb{X}_1 \times \cdots \times \mathbb{X}_N$.

Theorem 4: The N-tuple $\check{\mathcal{U}}$, formed by $\check{\mathcal{U}}_i = \mathcal{V}^{\bar{1}}_{X_i}$ for all i in $\{1, \dots, N\}$, minimizes $\mathcal{J}_B$.

Proof: The fact that $\hat{W}$ is a MAP estimator leads to the following equalities:

$$\mathbf{P}(W = \hat{W} | Y = \varnothing) = \prod_{k=1}^{N} \max_{\alpha_k \in \mathbb{X}_k} \mathbf{P}(X_k = \alpha_k | U_k = 0) \tag{123}$$

and

$$\mathbf{P}(W = \hat{W} | Y = (j, \tilde{x})) = \prod_{k \neq j} \max_{\alpha_k \in \mathbb{X}_k} \mathbf{P}(X_k = \alpha_k | U_k = 0). \tag{124}$$

The fact that the estimator receives U from $\aleph$ when a collision occurs, leads to:

$$\mathbf{P}(W = \hat{W} | Y = (\mathfrak{C}, \mu)) = \prod_{k=1}^{N} \max_{\alpha_k \in \mathbb{X}_k} \mathbf{P}(X_k = \alpha_k | U_k = \mu_k). \tag{125}$$

for any μ in $\{0,1\}^N$. For any given policy $\mathcal{U}$, we can use the total probability law to express the cost as:

$$\begin{aligned}\mathcal{J}_B(\mathcal{U}) = 1 &- \prod_{k=1}^{N} \max_{\alpha_k \in \mathbb{X}_k} \mathbf{P}(X_k = \alpha_k, U_k = 0) \\ &- \sum_{j=1}^{N} \Big(\prod_{k \neq j} \max_{\alpha_k \in \mathbb{X}_k} \mathbf{P}(X_k = \alpha_k, U_k = 0)\Big) \mathbf{P}(U_j = 1) \\ &- \sum_{\mu \in \mathbb{L}_{N,2}} \Big(\prod_{k=1}^{N} \max_{\alpha_k \in \mathbb{X}_k} \mathbf{P}(X_k = \alpha_k, U_k = \mu_k)\Big). \end{aligned} \tag{126}$$

where $\mathbb{L}_{N,k} \stackrel{\text{def}}{=} \{\mu \in \{0,1\}^N | \sum_{i=1}^{N} \mu_i \geq k\}$.

The cost can be re-expressed as:

$$\begin{aligned}\mathcal{J}_B(\mathcal{U}) = 1 &- \prod_{k=1}^{N} \max_{\alpha_k\in\mathbb{X}_k}(1-\mathcal{U}_k(\alpha_k))p_{X_k}(\alpha_k)\\ &- \sum_{j=1}^{N}\Big(\prod_{k\neq j}\max_{\alpha_k\in\mathbb{X}_k}(1-\mathcal{U}_k(\alpha_k))p_{X_k}(\alpha_k)\Big)\\ &\qquad\times\sum_{\alpha_j\in\mathbb{X}_j}\mathcal{U}_j(\alpha_j)p_{X_j}(\alpha_j)\\ &- \sum_{\mu\in\mathbb{L}_{N,2}}\Big(\prod_{k:\mu_k=0}\max_{\alpha_k\in\mathbb{X}_k}(1-\mathcal{U}_k(\alpha_k))p_{X_k}(\alpha_k)\Big)\\ &\qquad\times\Big(\prod_{k:\mu_k=1}\max_{\alpha_k\in\mathbb{X}_k}\mathcal{U}_k(\alpha_k)p_{X_k}(\alpha_k)\Big). \qquad (127)\end{aligned}$$

For any arbitrary choice of i in $\{1,\dots,N\}$, and for any given fixed $\tilde{\mathcal{U}}_{-i}$, we have:

$$\begin{aligned}\mathcal{J}_B(\mathcal{U}_i,\tilde{\mathcal{U}}_{-i}) = 1 &- \tilde{\tau}_{-i}\max_{\alpha_i\in\mathbb{X}_i}\mathcal{U}_i(\alpha_i)p_{X_i}(\alpha_i)\\ &- \tilde{\varrho}_{-i}\sum_{\alpha_i\in\mathbb{X}_i}\mathcal{U}_i(\alpha_i)p_{X_i}(\alpha_i)\\ &- (\tilde{\varrho}_{-i}+\tilde{\beta}_{-i})\max_{\alpha_i\in\mathbb{X}_i}(1-\mathcal{U}_i(\alpha_i))p_{X_i}(\alpha_i), \qquad (128)\end{aligned}$$

where the coefficients $\tilde{\varrho}_{-i}, \tilde{\tau}_{-i}$ and $\tilde{\beta}_{-i}$ are given by:

$$\tilde{\varrho}_{-i} \overset{\text{def}}{=} \prod_{k\neq i}\max_{\alpha_k\in\mathbb{X}_k}(1-\tilde{\mathcal{U}}_k(\alpha_k))p_{X_k}(\alpha_k); \qquad (129)$$

$$\begin{aligned}\tilde{\tau}_{-i} \overset{\text{def}}{=} \sum_{\mu_{-i}\in\mathbb{L}_{N-1,1}}&\left(\prod_{k:\mu_k=0}\max_{\alpha_k\in\mathbb{X}_k}(1-\tilde{\mathcal{U}}_k(\alpha_k))p_{X_k}(\alpha_k)\right)\\ &\times\left(\prod_{k:\mu_k=1}\max_{\alpha_k\in\mathbb{X}_k}\tilde{\mathcal{U}}_k(\alpha_k)p_{X_k}(\alpha_k)\right); \qquad (130)\end{aligned}$$

and

$$\begin{aligned}\tilde{\beta}_{-i} \overset{\text{def}}{=} \sum_{\mu_{-i}\in\mathbb{L}_{N-1,2}}&\left(\prod_{\substack{k\neq i:\\ u_k=0}}\max_{\alpha_k\in\mathbb{X}_k}(1-\tilde{\mathcal{U}}_k(\alpha_k))p_{X_k}(\alpha_k)\right)\\ &\times\left(\prod_{\substack{k\neq i:\\ \mu_k=1}}\max_{\alpha_k\in\mathbb{X}_k}\tilde{\mathcal{U}}_k(\alpha_k)p_{X_k}(\alpha_k)\right)\\ &+\sum_{j\neq i}\left(\prod_{k\notin\{i,j\}}\max_{\alpha_k\in\mathbb{X}_k}(1-\tilde{\mathcal{U}}_k(\alpha_k))p_{X_k}(\alpha_k)\right)\\ &\qquad\times\sum_{\alpha_j\in\mathbb{X}_j}\tilde{\mathcal{U}}_j(\alpha_j)p_{X_j}(\alpha_j). \qquad (131)\end{aligned}$$

After a few calculations, we arrive at the following:

$$\begin{aligned}\tilde{\tau}_{-i}-\tilde{\beta}_{-i} = \sum_{j\neq i}&\left(\prod_{k\notin\{i,j\}}\max_{\alpha_k\in\mathbb{X}_k}(1-\tilde{\mathcal{U}}_k(\alpha_k))p_{X_k}(\alpha_k)\right)\\ \times&\left(\max_{\alpha_j\in\mathbb{X}_j}\tilde{\mathcal{U}}_j(\alpha_j)p_{X_j}(\alpha_j)-\sum_{\alpha_j\in\mathbb{X}_j}\tilde{\mathcal{U}}_j(\alpha_j)p_{X_j}(\alpha_j)\right)\\ &\le 0. \qquad (132)\end{aligned}$$

From the inequality above, which implies that $\tilde{\tau}_{-i}\le\tilde{\beta}_{-i}$, and from the fact that if $\tilde{\mathcal{U}}_{-i}$ is fixed then Eq. (128) has the same structure as $\mathscr{J}_B$ in Eq. (102), we can conclude from Lemma 4 that $\mathcal{J}_B(\mathcal{V}^{\bar{1}}_{X_i},\tilde{\mathcal{U}}_{-i})\le\mathcal{J}_B(\mathcal{U}_i,\tilde{\mathcal{U}}_{-i})$ holds for any $\mathcal{U}_i$ in $\mathbb{U}_i$. Since, the choice of i and $\tilde{\mathcal{U}}_{-i}$ was arbitrary, we conclude that $\check{\mathcal{U}}$, formed by $\check{\mathcal{U}}_i=\mathcal{V}^{\bar{1}}_{X_i}$ for all i in $\{1,\dots,N\}$, is globally optimal. ■

C. Illustrative example

Consider the case of a system where N sensors observe independent random variables $\{X_1,\dots,X_N\}$ that are identically distributed according to p_X. The cost evaluated at a globally optimal solution $\mathcal{U}^\star=(\mathcal{V}^{\bar{1}}_{X_1},\dots,\mathcal{V}^{\bar{1}}_{X_N})$ gives:

$$\begin{aligned}\mathcal{J}_B(\mathcal{U}^\star) =\; &1-N(q_X(1))^{N-1}(1-q_X(1)-q_X(2))\\ &-(q_X(1)+q_X(2))^N. \qquad (133)\end{aligned}$$

When p_X is Bernoulli, it follows immediately that the optimal cost is zero for any number of sensors. However, for $|\mathbb{X}|\ge 3$, the performance of the system degrades when the number of sensors N increases and the optimal cost converges to one as N tends to infinity.

We illustrate how the performance degrades with N for the case in which p_X is the geometric probability mass function with parameter $\pi\in[0,1]$. The optimal cost calculated for $N=2,4,8,16$ and 64 is depicted in Fig. 5.

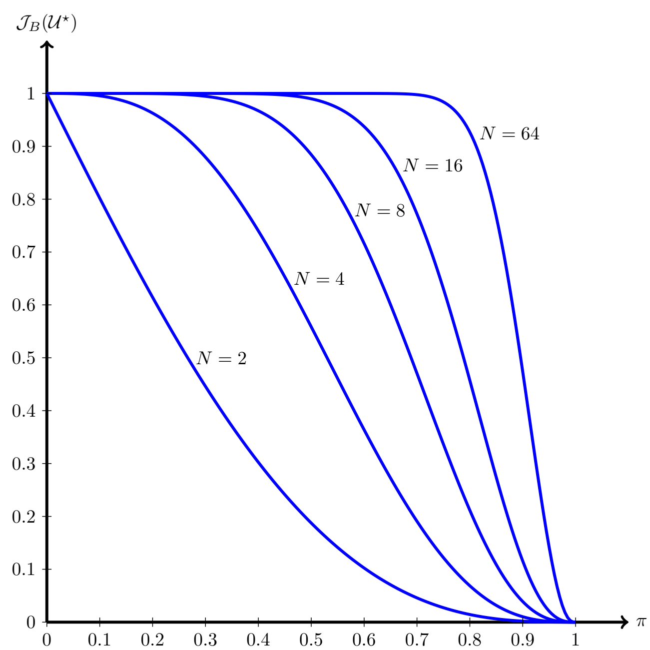

Fig. 5. Optimal performance of a team with N sensors observing i.i.d. Geometric random variables with parameter π and minimizing the total probability of error criterion.

VI. Conclusion and future work

In this paper, we propose a framework in which estimates of two or more independent discrete random variables are calculated based on information that is transmitted via a collision channel. Each random variable is observed by a sensor that decides whether it must stay silent or attempt to transmit its measurement to the estimator. The collision channel declares a collision when there are two or more transmission attempts, it also informs the estimator when all sensors choose to be silent and, otherwise, it transmits unerringly the value of a measurement the estimator.

A team decision problem is formulated that seeks to determine the transmission policies that minimize two types of cost. The first is a weighted sum of the probabilities of error of the estimates of each random variable. We also considered the case in which we seek to minimize the total probability of error. In both cases, we used person-by-person optimality principles to obtain an optimal solution. For the latter case, we explicitly characterize an optimal solution, while in the former we show that there is a finite set of candidate solutions that contain at least an optimal one. The number of candidate solutions grows exponentially (3^N) with the number of sensors but it does not depend on the distribution or cardinality of the support of the random variables, which can be countably infinite. The determination of which solution is optimal among the set of candidates, as well as the determination of the minimal cost, is carried out via simple calculations that depend only on the probabilities of the two most likely outcomes of each random variable. When there are two sensors, we show that the number of candidate solutions can be further reduced from nine to five. This suggests that 3^N may be, in general, a very conservative upper bound on the number of candidate solutions that need to be tested.

Open problems: There are several possible extensions of the work reported here. First, is to consider the case when the random variables are dependent or, as addressed in [27] for continuously distributed random variables, the measurements contain private and common components. The case in which u is not available in Eq. (113) when there is a collision is a realistic and important problem that also remains unsolved. In this case, there is ambiguity on which sensors attempted to transmit when a collision occurs. It is also important to investigate systematic methods to reduce the number of candidate solutions that need to be tested when the cost is the weighted sum of the probabilities of error. Finally, the sequential estimation of discrete Markov sources over the collision channel with feedback is yet another important unsolved problem.

References

[1] P. J. Antsaklis, "Goals and challenges in cyber-physical systems research," *IEEE Transactions on Automatic Control*, vol. 59, no. 12, pp. 3117–3119, December 2014.

[2] I. F. Akyildiz, F. Brunetti, and C. Blázquez, "Nanonetworks: A new communication paradigm," *Computer Networks*, vol. 52, no. 12, pp. 2260 – 2279, August 2008.

[3] B. Atakan and O. B. Akan, "Carbon nanotube-based nanoscale ad hoc networks," *IEEE Communications Magazine*, vol. 48, no. 6, pp. 129 – 135, June 2010.

[4] K.-D. Kim and P. R. Kumar, "Cyber-physical systems: A perspective at the centennial," *Proceedings of the IEEE*, vol. 100, no. 13, pp. 1287–1308, May 2012.

[5] K. Savla, G. Como, and M. A. Dahleh, "Robust network routing under cascading failures," *IEEE Transactions on Network Science and Engineering*, vol. 1, no. 1, pp. 53–66, January–June 2014.

[6] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715 – 2729, November 2013.

[7] S. Yuksel and T. Basar, *Stochastic Networked Control Systems: Stabilization and Optimization Under Information Constraints*. Birkhauser, 2013.

[8] J. Tsitsiklis and M. Athans, "On the complexity of decentralized decision making and detection problems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 440–446, May 1985.

[9] R. Cogill and S. Lall, "An approximation algorithm for the discrete team decision problem," *SIAM Journal on Control and Optimization*, vol. 45, no. 4, pp. 1359–1368, 2006.

[10] P. R. de Waal and J. H. van Schuppen, "A class of team problems with discrete action spaces: Optimality conditions based on multimodularity," *SIAM Journal on Control and Optimization*, vol. 38, no. 3, pp. 875 – 892, 2000.

[11] G. C. Goodwin and D. Quevedo, "Finite alphabet control and estimation," *International Journal of Control, Automation, and Systems*, vol. 1, no. 4, pp. 412 – 430, December 2003.

[12] H. S. Witsenhausen, "On the structure of real-time source coders," *The Bell System Technical Journal*, vol. 58, no. 6, pp. 1437 – 1451, July – August 1978.

[13] J. C. Walrand and P. Varaiya, "Optimal causal coding - decoding problems," *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 814 – 820, November 1983.

[14] D. Teneketzis, "On the structure of optimal real-time encoders and decoders in noisy communication," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4017 – 4035, September 2006.

[15] A. Mahajan and D. Teneketzis, "Optimal design of sequential real-time communication systems," *IEEE Transactions on Information Theory*, vol. 55, no. 11, pp. 5317 – 5338, November 2009.

[16] O. C. Imer and T. Basar, "Optimal estimation with limited measurements," *International Journal on Systems, Control and Communications*, vol. 2, no. 1/2/3, pp. 5 – 29, 2010.

[17] G. M. Lipsa and N. C. Martins, "Remote state estimation with communication costs for first-order lti systems," *IEEE Transactions on Automatic Control*, vol. 56, no. 9, pp. 2013–2025, September 2011.

[18] A. Nayyar and D. Teneketzis, "On the structure of real-time encoders and decoders in a multi-terminal communication system," *IEEE Transactions on Information Theory*, vol. 57, no. 9, pp. 6196–6214, September 2011.

[19] M. M. Vasconcelos and N. C. Martins, "Optimal estimation over the collision channel," *IEEE Transactions on Automatic Control*, vol. 62, no. 1, pp. 321–336, January 2017.

[20] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 2006.

[21] S. Adlakha, S. R. Mirghaderi, S. Lall, and A. Goldsmith, "Decentralized team decision via coding," in *47th Annual Allerton Conference on Communication, Control and Computing*, 2009, p. 151.

[22] E. J. Msechu and G. B. Giannakis, "Sensor-centric data reduction for estimation with WSNs via censoring and quantization," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 400 – 414, January 2012.

[23] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 49, no. 9, pp. 1453–1464, 2004.

[24] S. Boyd and L. Vandenbergue, *Convex Optimization*. New York, NY, USA: Cambridge Univeristy Press, 2004.

[25] D. G. Luenberger, *Optimization by vector space methods*. Wiley-Interscience, 1969.

[26] K. Whitehouse, A. Woo, F. Jiang, J. Polastre, and D. Culler, "Exploiting the capture effect for collision detection and recovery," in *Proceedings of the 2nd IEEE workshop on Embedded Networked Sensors*, 2005, pp. 45 – 52.

[27] M. M. Vasconcelos and N. C. Martins, "The structure of optimal communication policies for remote estimation over the collision channel with private and common observations," in *Proceedings of the 55th IEEE Conference on Decision and Control*, Las Vegas, NV, December 2016, pp. 320–326.