ISIT 2009, Seoul, Korea, June 28 - July 3, 2009

Distributed Lossy Averaging

Han-I Su
Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA
Email: hanisu@stanford.edu

Abbas El Gamal
Department of Electrical Engineering
Stanford University
Stanford, CA 94305, USA
Email: abbas@stanford.edu

***Abstract*—An information theoretic formulation of distributed averaging is presented. We assume a network with m nodes each observing an i.i.d. source; the nodes communicate and perform local processing with the goal of computing the average of the sources to within a prescribed mean squared error distortion. The network rate distortion function $R^*(D)$ for a 2-node network with correlated Gaussian sources is established. A general cutset lower bound on $R^*(D)$ with independent Gaussian sources is established and shown to be achievable to within a factor of 2 via a centralized protocol. A lower bound on the network rate distortion function for distributed weighted-sum protocols that is larger than the cutset bound by a factor of $\log m$ is established. An upper bound on the expected network rate distortion function for gossip-based weighted-sum protocols that is only a factor of $\log \log m$ larger than this lower bound is established. The results suggest that using distributed protocols results in a factor of $\log m$ increase in communication relative to centralized protocols.**

I. Introduction

Distributed averaging is a popular example of the distributed consensus problem, which has been receiving much attention recently due to interest in applications ranging from distributed coordination of autonomous agents to distributed computation in sensor networks, ad-hoc networks, and peer-to-peer networks.

This paper presents a lossy source coding formulation of the distributed averaging problem. We assume that each node in the network observes a source and the nodes communicate and perform local processing with the goal of computing the average of the sources to within a prescribed mean squared error distortion. We investigate the network rate distortion function in general and for the class of weighted-sum protocols, including random gossip-based protocols.

Most previous work on distributed averaging, e.g., [1], [2], has involved the noiseless communication and computation of real numbers, which is unrealistic. Recognizing this shortcoming, the effect of quantization on distributed averaging has been recently investigated. Our work is related most closely to the work in [3]–[5]. Compared to [3], [4], our information-theoretic approach deals more naturally and fundamentally with quantization and provides limits that hold independent of implementation details. Our results, however, cannot be compared directly to results in these papers because of differences in the models and assumptions. While the work in [5] is information-theoretic, it deals with a different formulation than ours and the results are not comparable. Our formulation of the distributed averaging problem can be viewed also as a generalization of the CEO problem [6], where in our setting every node wishes to compute the average and the communication protocol is significantly more complex in that it allows for interactivity, relaying, and local computing, in addition to multiple access.

In the following section, we introduce the lossy averaging problem. In Section III, we establish the network rate distortion function for a 2-node network. In Section IV, we establish a general cutset lower bound on the network rate distortion function and show that it can be achieved within a factor of 2 using a centralized protocol. In Section V, we investigate the class of distributed weighted-sum protocols. We establish a lower bound on the network rate distortion function for this class as well as an upper bound for gossip-based weighted-sum protocols. The full paper is posted on arXiv.org [7].

II. Lossy Averaging Problem

Consider a network with m sender-receiver nodes, where node $i = 1, 2, \ldots, m$ observes an i.i.d. source X_i. The nodes communicate and perform local processing with the goal of computing the average of the sources $S = (1/m) \sum_{i=1}^m X_i$ at each node to within a prescribed distortion D. The following definitions apply to any set of correlated sources $(X_1, X_2, \ldots, X_m)$. In Sections IV and V, we assume that the sources are independent white Gaussian noise (WGN) processes each with average power of one.

The topology of the network is specified by a connected graph $(\mathcal{M}, \mathcal{E})$ without self loops, where $\mathcal{M} = \{1, 2, \ldots, m\}$ is the set of nodes and $\mathcal{E}$ is a set of undirected edges (node pairs) $\{i, j\}$, $i, j \in \mathcal{M}$ for $i \neq j$. Communication is performed in *rounds* and each round is divided into *time slots*. Each round may consist of a different number of time slots, and each time slot may consist of a different number of transmissions. One edge (node pair) is chosen at each round and only one node is allowed to transmit in each time slot. Without loss of generality we assume that the selected node pair communicate in a round robin manner with the first node communicating in odd time slots and the second node communicating in even time slots. Further, we assume a *source coding* setting, where communication is noiseless and instant, that is, every transmitted message is successfully received by the intended receiver in the same time slot it is transmitted in.

Communication and computing are performed according to an agreed upon *averaging protocol* that determines (i) the