Signal Processing 93 (2013) 109–123

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Approximate decoding approaches for network coded correlated data

Hyunggon Park [a,*], Nikolaos Thomos [b,c], Pascal Frossard [b]

[a] Multimedia Communications and Networking Laboratory, Ewha Womans University, Seoul, Republic of Korea
[b] Signal Processing Lab. (LTS4), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
[c] Communication and Distributed Systems laboratory (CDS), University of Bern, Bern, Switzerland

ARTICLE INFO

Article history:
Received 30 December 2011
Received in revised form
19 June 2012
Accepted 8 July 2012
Available online 20 July 2012

Keywords:
Network coding
Approximate decoding
Correlated data
Distributed transmission
Ad hoc networks

ABSTRACT

This paper considers a framework where data from correlated sources are transmitted with the help of network coding in ad hoc network topologies. The correlated data are encoded independently at sensors and network coding is employed in the intermediate nodes in order to improve the data delivery performance. In such settings, we focus on the problem of reconstructing the sources at decoder when perfect decoding is not possible due to losses or bandwidth variations. We show that the source data similarity can be used at decoder to permit decoding based on a novel and simple approximate decoding scheme. We analyze the influence of the network coding parameters and in particular the size of finite coding fields on the decoding performance. We further determine the optimal field size that maximizes the expected decoding performance as a trade-off between information loss incurred by limiting the resolution of the source data and the error probability in the reconstructed data. Moreover, we show that the performance of the approximate decoding improves when the accuracy of the source model increases even with simple approximate decoding techniques. We provide illustrative examples showing how the proposed algorithm can be deployed in sensor networks and distributed imaging applications.

1. Introduction

The rapid deployment of distributed networks such as sensor networks, cloud networks has motivated a plethora of researches that study the design of low complexity and efficient solutions for information delivery. Since the coordination among intermediate nodes is often difficult to achieve, the information dissemination in the intermediate nodes has often to be performed in a distributed manner on ad hoc or overlay mesh network topologies. Network coding [1] has been recently proposed as a method to build efficient distributed delivery algorithms in networks with path and source diversity. It is based on a paradigm where the network nodes are allowed to perform basic processing operations on information streams. The network nodes can combine information packets and transmit the resulting data to the next network nodes. Such a strategy permits to improve the throughput of the system and to approach better the max-flow min-cut limit of networks [2,3]. When the decoder receives enough data, it can recover the original source information by performing inverse operations (e.g., with Gaussian elimination).

These advantages motivate the deployment of network coding in various scenarios where the network diversity is significant (e.g., [4–9]). Many of these solutions are based on random linear network coding (RLNC) [10] that permits to implement distributed solutions with low communication costs. RLNC represents an interesting solution for the deployment of practical systems where it can work in

* Corresponding author.
E-mail address: hyunggon.park@ewha.ac.kr (H. Park).

0165-1684/$ - see front matter
http://dx.doi.org/10.1016/j.sigpro.2012.07.007

conjunction with data dissemination protocols such as gossiping algorithms [8]. The resulting systems are robust against link failures, do not require reconciliation between the network nodes, and can significantly improve the performance of data delivery compared to “store and forward” approaches. Most of the research so far has however focused either on theoretical aspects of network coding such as achievable capacity and coding gain or on its practical aspects such as robustness and increased throughput when the number of innovative packets is sufficient for perfect decoding. However, it generally does not consider the problematic cases where the clients receive an insufficient number of innovative packets for perfect decoding due to losses or timing constraints for example. This is the main problem addressed in this paper.

We consider a framework where network coding is used for the delivery of correlated data that are discretized and independently encoded at the sensors. The information streams are delivered with the help of RLNC in lossy ad hoc networks. When an insufficient number of symbols at decoder prevent exact data recovery, we design a novel low complexity *approximate decoding* algorithm that uses the data correlation for signal reconstruction. The information about source similarity typically provides additional constraints in the decoding process, such that well-known approaches for matrix inversion (e.g., Gaussian elimination) can be efficiently used even in the case where the decoding problem is a priori underdetermined. We show analytically that the use of source models at decoding process leads to an improved data recovery. Then, we analyze the impact of accurate knowledge of data similarity at decoder, where more precise information leads to better performance in the approximate decoding. We further analyze the influence of the choice of the Galois Field (GF) size in the coding operations on the performance of the approximate decoding framework. We demonstrate that the field size should be selected by considering the tradeoff between the resolution for representing the source signal and the approximate decoding performance. Specifically, when the GF size increases, the quantization error of the source data decreases, while the decoding error probability increases with the GF size. We show that there is an optimal value for the GF size when the approximate decoding is enabled at the receivers. Finally, we illustrate the performance of the network coding algorithm with the approximate decoding on two types of correlated data, i.e., seismic data and video sequences. The simulation results confirm the validity of the GF size analysis and show that the approximate decoding scheme leads to efficient reconstruction when the accurate correlation information is used during decoding. In summary, the main contributions of our paper are (i) a new framework for the distributed delivery of correlated data with network coding, (ii) a novel approximate decoding strategy that exploits data similarity with low complexity when the received data does not permit perfect decoding, (iii) an analysis of the influence of the accuracy of the data similarity information and the GF size on the decoding performance, and (iv) the implementation of illustrative examples with external or intrinsic source correlation.

In general, the transmission of correlated sources is studied in the framework of distributed coding [11] (i.e., in the context of the Slepian–Wolf problem), where sources are typically encoded by systematic channel encoders and eventually decoded jointly [12,13]. DSC (distributed source coding) is combined with network coding schemes [14–17] in the gathering of correlated data. Alternatively, network coding techniques are used while jointly performing data compression [18,19]. Our focus is however not on the design of a distributed compression scheme, which generally assumes that sensors are aware of the similarity between the data sources. Rather, we focus on the transmission of correlated data that are encoded independently, transmitted with the help of network coding enabled network nodes over an overlay network and jointly decoded at the receivers. However, due to the network dynamics, there is no guarantee that each node receives enough useful packets for successful data recovery. Hence, it is essential to have a low complexity methodology that enables the recovery of the original data with a good accuracy, when the number of useful packets is not sufficient for perfect decoding. When RLNC is implemented in the network, the encoding processes in each node are based on linear operations (e.g., linear combinations, inverse of linear matrix, etc.) in a finite algebraic field. In the case of insufficient number of innovative packets for perfect decoding, one can simply deploy an existing regularization technique that may minimize the norm of the errors using the pseudo-inverse of the encoding matrix. However, it is generally known that this type of regularization techniques may result in significantly unreasonable approximation [20]. Alternatively, Tikhonov regularization provides an improved performance by slightly modifying the standard least square formula. However, this technique requires to determine additional optimization parameters, which is nontrivial in practice. Sparsity assumptions might also be used [21] for regularized decoding in underdetermined systems in cases where a model of the signal of interest is known a priori. However, all these regularization techniques have been designed and developed in the continuous domain, but not for finite fields that are used in network coding approaches. Thus, they may show significantly poor performance if they are applied blindly in our framework, as they cannot consider several properties (e.g., cyclic properties) of finite field operations. Underdetermined systems can also be solved approximately based on the maximum likelihood estimation (MLE) techniques (see e.g., [22] (Part II)) or based on mixed integer linear programming [23], but these techniques require effective data models and typically involve large computational complexity.

The paper is organized as follows. In Section 2, we present our framework and describe the approximate decoding algorithm. We discuss the influence of the source model information in the approximate decoding process in Section 3. In Section 4, we analyze the relation between the decoding performance and the GF size, and then determine an optimal GF size that achieves the smallest expected decoding error. Sections 5 and 6 provide illustrative examples that show how the proposed approach can be implemented in sensor networks or video delivery applications.