Signal Processing 120 (2016) 222–235

Contents lists available at ScienceDirect

Signal Processing

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

Approximate decoding for network coded inter-dependent data

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

[a] Department of Electronics Engineering, Ewha Womans University, Seoul, Republic of Korea
[b] Computer Science and Electronic Engineering Department, University of Essex, Colchester, United Kingdom
[c] Signal Processing Laboratory (LTS4), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 17 December 2014
Received in revised form
12 July 2015
Accepted 14 September 2015
Available online 26 September 2015

Keywords:
Network coding
Approximate decoding
Inter-dependence source
Ad hoc networks
Loss-tolerant data

ABSTRACT

In this paper, we consider decoding of loss tolerant data encoded by network coding and transmitted over error-prone networks. Intermediate network nodes typically perform the random linear network coding in a Galois field and a Gaussian elimination is used for decoding process in the terminal nodes. In such settings, conventional decoding approaches can unfortunately not reconstruct any encoded data unless they receive at least as many coded packets as the original number of packets. In this paper, we rather propose to exploit the incomplete data at a receiver without major modifications to the conventional decoding architecture. We study the problem of approximate decoding for inter-dependent sources where the difference between source vectors is characterized by a unimodal distribution. We propose a mode-based algorithm for approximate decoding, where the *mode* of the source data distribution is used to reconstruct source data. We further improve the mode-based approximate decoding algorithm by using additional short information that is referred to as position similarity information (PSI). We analytically study the impact of PSI size on the approximate decoding performance and show that the optimal size of PSI can be determined based on performance requirements of applications. The proposed approach has been tested in an illustrative example of data collection in sensor networks. The simulation results confirm the benefits of approximate decoding for inter-dependent sources and further show that 93.3% of decoding errors are eliminated when the approximate decoding uses appropriate PSI.

1. Introduction

The hardware development of sensors and communication chipsets has enabled easy deployment of sensor networks and it has led to an excessive network traffic and demands to increase network capacity. Network coding [1] has been proposed in order to increase the throughput of networks; it can reach the max-flow capacity between the source and each destination node [2–4]. In this case, unlike simple data forwarding in conventional networks, intermediate network nodes combine the received packets with basic coding operations. Network coding can lead to efficient resources usage (e.g., bandwidth and power), reduced computations, and improved robustness against network dynamics [5] by exploiting the diversity in networks. A variety of applications have been developed by taking advantages of network coding (e.g., content distribution, storages, and P2P systems [6–10]). Random linear network coding (RLNC) [11] is the most popular network coding algorithm, as it permits distributed deployment in dynamic error-prone networks [12].

* Corresponding author.
E-mail addresses: minhae.kwon@ewhain.net (M. Kwon), hyunggon.park@ewha.ac.kr (H. Park), nthomos@essex.ac.uk (N. Thomos), pascal.frossard@epfl.ch (P. Frossard).

http://dx.doi.org/10.1016/j.sigpro.2015.09.010

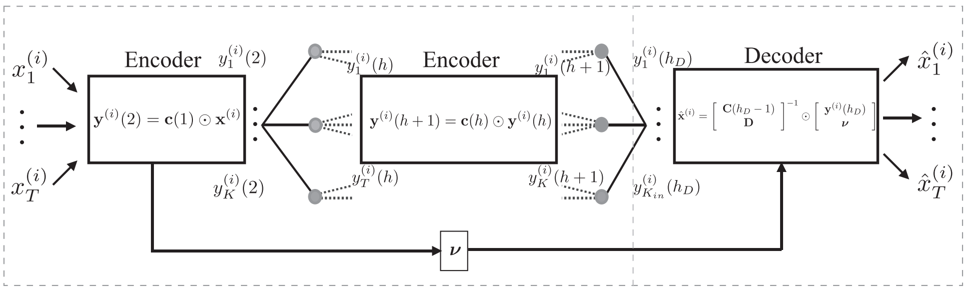

Fig. 1. Overall description of system setup. At the h-th coding stage, the incoming T innovative packets $y_m^{(i)}(h)$ $(1 \le m \le T)$ are combined based on RLNC and K outgoing packets are generated. If T innovative packets are not available for the decoder at the moment of decoding (i.e., $K_{in} < T$), the proposed approximate decoding is deployed with side information $\boldsymbol{\nu}$, which is delivered from the encoder. These are discussed in Sections 2.2 and 2.3, respectively.

While several advantages can be obtained by deploying network coding techniques for information delivery, it has a significant drawback in practice, which is also known as *all-or-nothing problem*, i.e., a terminal node cannot recover any information from the received data unless it receives at least the same number of innovative coded packets[1] as the number of source packets. In other words, under the conventional decoding process (e.g., Gaussian elimination), the received packets have to form a full-rank system for decoding. However, perfect decoding might not always be necessary and approximate reconstruction may be sufficient for several services that can accept imperfect reconstruction.

In order to solve this problem, we propose an approach to approximately recover inter-dependent sources from a set of network coded data that does not form a full rank system at receiver. With this same objective, a low complexity approximate decoding algorithm has been presented in [14], where the receiver simply matches the most similar data between neighbor sources and thus reaches only limited approximate decoding performance. In this paper, we present an improved approximate decoding algorithm that exploits the source characteristics, i.e., the distribution of differences between neighbor source vectors, thereby explicitly considering more general types of source data. We propose to use the *mode* of the distribution (i.e., the value that appears most often in a distribution) in the source characteristics to build an approximate decoding algorithm. The mode of the distribution is referred to as *similarity information* (SI). We show that it is sufficient side information to maximize performance of the proposed approximate decoding. As a result, the mode-based approximate decoding can significantly reduce the amount of side information needed for decoding. The decoding performance can be further improved by considering the positions where errors may occur, which are explicitly captured by the *position similarity information* (PSI) at the expense of additional side information. We investigate the tradeoff between the PSI size (i.e., the amount of side information or communication overheads) and the corresponding decoding performance and show that there is an optimal amount of additional information for approximate decoding. Finally, the proposed approach is deployed in an illustrative example of sensor networks and the simulation results confirm our theoretical performance study.

The main contributions of the paper can be summarized as follows:

- we propose a generalized framework of approximate decoding that covers large range of source types,
- we develop an algorithm that enables the approximate decoding solution to be deployed for any linearly interdependent sources,
- we develop decoding algorithms that can exploit both SI and PSI, leading to significantly improved decoding performance,
- we analytically study the tradeoff between communication overhead (incurred by deploying SI and PSI) and decoding performance gains, and
- we have extensive set of experiment results that confirm the theoretical analysis.

The rest of the paper is organized as follows. The general network coding framework is presented in Section 2. The mode-based decoding approximate decoding algorithm that considers inter-dependent source distributions is proposed and discussed in Section 3. In Section 4, we show that the decoding performance can be improved by incorporating PSI into the mode-based approach. In Section 5, we evaluate and compare the performance of the mode-based approach against conventional decoding methods in an illustrative sensor network scenario. Related works are discussed in Section 6 and conclusion is drawn in Section 7.

2. System setup

We consider data transmission over error-prone networks that consist of source nodes, intermediate nodes and receivers. The source data is delivered to the receivers through intermediate nodes that are able to perform network coding, similar to the frameworks in [15–17]. The overview of the proposed system is shown in Fig. 1 and the details will be discussed next.

[1] A packet is *innovative* for a node if its coding vector is not in the span of the coding vectors of the packets already available at the node [13].