Computers and Electrical Engineering 45 (2015) 55–67

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Systematic rateless erasure code for short messages transmission ☆

Zan-Kai Chong [a,*], Bok-Min Goi [a], Hiroyuki Ohsaki [c], Bryan Cheng-Kuan Ng [d], Hong-Tat Ewe [b]

[a] Department of Electrical and Electronic Engineering, Universiti Tunku Abdul Rahman, Bandar Sungai Long, Cheras, 43000 Kajang, Selangor Darul Ehsan, Malaysia
[b] Department of Internet Engineering and Computer Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, Cheras, 43000 Kajang, Selangor Darul Ehsan, Malaysia
[c] Department of Informatics School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda, Hyogo 669-1337, Japan
[d] Orange Labs, 905 rue Albert Einstein, Sophia Antipolis, 06921, Antibes, Nice, France

A R T I C L E I N F O

Article history:
Received 27 February 2014
Received in revised form 2 June 2015
Accepted 2 June 2015
Available online 24 June 2015

Keywords:
Channel coding
Rateless erasure code
Short messages transmission
Networking

A B S T R A C T

In this paper, we propose a systematic rateless erasure code, namely systematic Random (SYSR) code based on Random code for short messages transmission. Given a message of k symbols, the sender will first send the message to the receiver as Part I coded symbols. The rest of coded symbols starting from $(k+1)$th onwards are termed as Part II coded symbols and they are generated by adding the message symbols randomly (XOR operation). The receiver reconstructs the original message instantly if all the Part I coded symbols are received intact. Otherwise, the receiver reconstructs the original message from any $k+10$ coded symbols of Part I and II with high probability of complete decoding (PCD), i.e. 99.9% success probability. Though SYSR code inherits the high decoding complexity of Random code, i.e. $O(k^3)$, both analysis and simulation results show that SYSR code achieves better PCD and fewer decoding steps than Random code.

1. Introduction

The rateless erasure code (also known as fountain code) is a class of erasure code, where the sender generates potentially infinite number of coded symbols from a message of k symbols. Upon receiving $k(1+\epsilon)$ coded symbols, the receiver reconstructs the original message with high probability of complete decoding (PCD), i.e. 99.9% success probability, where ϵ denotes the decoding inefficiency and $\epsilon \geqslant 0$.

To improve the decoding performance, the *systematic* rateless erasure code uses the k symbols of the original message as the first part of the coded symbols and the rest are generated using a sequence of bit operations on the original message. Then, the receiver reconstructs the original message instantly if the first k coded symbols are received intact. In case of any lost symbols, the receiver will reconstruct the original message from the subsequent coded symbols.

Significant research effort has been channelled to study rateless erasure codes in data networks, such as the Internet [1,2], vehicular ad-hoc network (VANET) [3,4], and delay-tolerant network (DTN) [5]. However, majority of the traffic in the Internet are short messages of less than 10 packets [6] and the VANET's messages are normally in a few hundred bytes

☆ Reviews processed and recommended for publication to the Editor-in-Chief by Associate Editor Dr. Fangyang Shen.

* Corresponding author.
E-mail addresses: chongzk@utar.edu.my (Z.-K. Chong), goibm@utar.edu.my (B.-M. Goi), ohsaki@kwansei.ac.jp (H. Ohsaki), cheng.kuan.ng@gmail.com (B.C.-K. Ng), eweht@utar.edu.my (H.-T. Ewe).

http://dx.doi.org/10.1016/j.compeleceng.2015.06.003
0045-7906/

[7]. The state-of-the-art rateless erasure codes (e.g. Luby Transform (LT) code [8] and Raptor code [9]) are inappropriate in the aforementioned networks due to their inefficiency in sending short messages. For example, the decoding inefficiency, ϵ is as high as 15% for the messages of thousands symbols [10] and it increases to about 30% when $k = 100$ [11]. Likewise, many non-systematic rateless erasure codes have been proposed for short messages transmission (cf. Section 2). These coding schemes require non-negligible decoding complexity even if all the coded symbols are received intact.

In this paper, we propose a systematic rateless erasure code, namely systematic Random code (SYSR code) that is built on top of random matrix (a matrix of randomly distributed binary values). Like typical systematic rateless erasure code, a message of k symbols will form part of the encoded symbols (termed as Part I coded symbols) and the rest of coded symbols from $k+1$th onwards (termed as Part II coded symbols) are generated by adding the k message symbols randomly, i.e. multiplying the random matrix with the message. Therefore, SYSR code is able to reconstruct the original message instantly if the Part I coded symbols are received intact. Otherwise, the original message can still be reconstructed with high PCD from any $k+10$ coded symbols of Part I and II with high PCD.

Unlike the rateless erasure codes that reviewed in Section 2, the high PCD of SYSR code is invariant to message length and applicable to short messages e.g. $k = 10$ symbols. Since random matrix is a widely used mathematical framework for coding theory, it is unsurprising that SYSR code looks similar to the network coding scheme in [12] at first glance. However, we point out that SYSR code is an erasure code that involves only sender and receiver whereas the network coding involves multiple network nodes cooperating in the encoding and decoding processes.

The remainder of the paper is structured as follows: The state-of-the-art rateless erasure codes and the properties of Random matrix to achieve full rank are discussed in Sections 2 and 3, respectively. Then, we propose SYSR code that is built on top of random matrix in Section 4 and the decoding process will be explained with an example. The PCD and decoding algorithm of SYSR code is analysed in Section 5 and the corresponding numerical results are presented in Section 6 alongside result with Random code and Stepping-Random (SR) code. Finally, we draw the conclusion in Section 7.

2. Related work

The erasure codes such as Reed-Solomon code is a fixed rate systematic erasure code [13]. Given a message of k symbols, the encoder will send the n coded symbols to the receiver, where the first k coded symbols belong to original message. Then, the receiver reconstructs the original message from any k out of n coded symbols, irrespective of the sequence and the code rate is k/n.

Generally, erasure codes require the sender to learn the channel erasure probability before initiating the encoding process. If more than $k - n$ coded symbols are lost in the erasure channel, there is no way for the receiver to reconstruct the original message. In this case, the sender needs to identify the lost packets and retransmit them manually. Such phenomenon underutilises the bandwidth.

To address the aforementioned issue, Byers et al. propose the idea of rateless erasure code [14]. Given a message of k symbols, the rateless erasure code generates potentially infinite number of coded symbols. Then, the receiver reconstructs the original message from any $k(1+\epsilon)$ coded symbols with high PCD, where ϵ denotes the decoding inefficiency. The sender does not require prior knowledge of the channel condition. It will keep sending the packets (i.e. coded symbols) until the receiver has sufficient packets (i.e. $k(1+\epsilon)$) to reconstruct the original message.

Luby Transform (LT) code [8] is the first practical rateless erasure code in the literature. A coded symbol is generated by selecting a degree d from the degree distribution (i.e. Soliton distribution) randomly. Then, d distinct message symbols are selected randomly and added together. The performance of LT code relies on the careful design of the degree distribution, which it ensures at least a degree one coded symbol exists in each iteration of the message passing algorithm during the decoding process. Then, Shokrollahi [9] further improves the decoding complexity of LT code by introducing the pre-code stage into the encoding process.

The aforementioned state-of-the-art rateless erasure codes are only efficient for long messages. To address the need in transmitting short messages, Hyytia et al. [15] and Zhang and Hranilovic [16] modify the degree distribution of LT code, respectively in order to improve the efficiency for short messages transmission. According to the published results, Hytia et al. require three to four extra coded symbols on average in order for a message of eight symbols to achieve complete decoding. On the other hand, the algorithm proposed by Zhang and Hranilovic [16] need extra 12 coded symbols to reconstruct a message of 32 symbols with high PCD. However, the aforementioned rateless erasure codes are not systematic in nature as well as those in [17–19]. For comparison, we will show that the proposed SYSR code achieves high PCD in $k+10$ coded symbols, or an extra of 1.6 coded symbols in average in Sections 5 and 6. On the contrary, the coding schemes in [20–22] are the systematic rateless erasure codes but they are not optimised for short messages transmission.

Random code is a rateless erasure code that uses a random matrix as the generator (see Section 3). It is able to achieve high PCD with $k+10$ coded symbols with the trade-off of high decoding complexity, i.e. $O\left(k^3\right)$ using Gaussian elimination. To the best of our understanding, the earliest discussion about Random code (properties of random matrix) is found in [23] and some relevant discussions about the PCD appear in [24,25].

Both Windowed code [10] and Stepping-Random (SR) code [26] are the non-systematic rateless erasure codes built on top of random matrix framework. Both of them reconstruct the original message from $k+10$ coded symbols with high PCD in