Raptor Codes

Amin Shokrollahi
EPFL and Digital Fountain, Inc.
Lausanne, Switzerland, and Fremont, USA
amin.shokrollahi@epfl.ch

Abstract — **We exhibit a class of universal Raptor Codes: for a given integer k, and any real $\varepsilon > 0$, Raptor Codes in this class produce a potentially infinite stream of symbols such that any subset of symbols of size $k(1+\varepsilon)$ is sufficient to recover the original k symbols, with high probability. Each output symbol is generated using $O(\log(1/\varepsilon))$ operations, and the original symbols are recovered from the collected ones with $O(k\log(1/\varepsilon))$ operations.**

I. Introduction

Raptor Codes are an extension of LT-Codes [1]. The parameters of a Raptor Code of length k over a field $\mathbb{F}$ are given by a pre-code $\mathcal{C}$ of dimension k and block-length n over $\mathbb{F}$, and a probability distribution Ω on $\mathbb{F}^n$. Given k source symbols $x_1, \ldots, x_k$, the pre-code first encodes these symbols into a codeword $(y_1, \ldots, y_n)$ of length n; the symbols belong to a finite dimensional vector space over $\mathbb{F}$. Each output symbol is obtained by sampling from the distribution Ω to obtain a vector $(v_1, \ldots, v_n)$. The value of the output symbol is then obtained as $\sum_{i=1}^n v_i y_i$.

The case of primary interest is when $\mathbb{F} = \mathbb{F}_2$, and Ω is a distribution which is constant on words of equal weight. In this case, Ω can be described by the numbers $\Omega_1, \ldots, \Omega_n$ such the probability of a vector $x \in \mathbb{F}^n$ under Ω is $\Omega_w / \binom{n}{w}$, where w is the Hamming weight of x. In this case, we identify Ω with the generating polynomial $\Omega(x) = \sum_{i=1}^n \Omega_i x^i$; the parameters of the Raptor Code then become $(k, \mathcal{C}, \Omega(x))$. Note that a Raptor Code does not have a fixed block-length. In applications, $x_1, \ldots, x_k$ can be packets to be sent over a computer network. The Raptor Code can be used to produce a potentially limitless stream of output symbols (i.e., packets). The design problem in this case consists of choosing the parameters of the Raptor Code in such a way that efficient decoding is possible after reception of $k(1+\varepsilon)$ output packets, for ε arbitrarily close to zero. We call such a decoding algorithm an algorithm *overhead* ε.

An encoding algorithm is called linear time if the pre-code can be encoded in linear time, and the average number of operations to produce an output symbol is a constant. A decoding algorithm is called linear time if, after collecting a certain number of output symbols, it can decode the k source symbols in time $O(k)$. We will solve the asymptotic design problem for Raptor Codes by showing that, for any given ε, an appropriate choice of the pre-code and the output degree distribution will result in the Belief Propagation (BP) Decoding algorithm to be linear time and of overhead ε.

II. Asymptotic Analysis

Let $\varepsilon > 0$, and let $\mathcal{C}_m$ be a sequence of codes of block-length n such that the rate R of $\mathcal{C}_m$ is (approximately) $(1+\varepsilon/2)/(1+\varepsilon)$, and such that the BP decoder can decode C_m on a BEC with erasure probability $\delta = (\varepsilon/4)/(1+\varepsilon) = (1-R)/2$ with $O(n\log(1/\varepsilon))$ arithmetic operations. Further, let $D := \lceil 4(1+\varepsilon)/\varepsilon \rceil$ and define

$$\Omega_D(x) = \frac{1}{\mu+1}\left(\mu x + \sum_{i=2}^{D} \frac{x^i}{i(i-1)} + \frac{x^{D+1}}{D}\right),$$

where $\mu = (\varepsilon/2) + (\varepsilon/2)^2$.

Theorem 1. *Let $R = (1+\varepsilon/2)/(1+\varepsilon)$, and C_m be the family of codes of rate R and dimension k with the properties described above. Then the Raptor code with parameters $(k, \mathcal{C}_n, \Omega_D(x))$ has a linear time encoding algorithm, and BP decoding is a linear time decoding of overhead ε for the code.* More precisely, it can be shown that the average number of operations to produce an output symbol is $O(\log(1/\varepsilon))$, and the average number of operations to recover the k source symbols is $O(k\log(1/\varepsilon))$. The theorem is proved using a combination of density evolution techniques from [2] and Luby's techniques [1].

III. Finite Length Design

We have developed efficient techniques to design Raptor Codes of small length for which the BP decoder has provable guarantees on its error performance. The design consists of a design problem for the output degree distribution $\Omega(x)$, and the design problem for the pre-code $\mathcal{C}$. For lack of space, we will not discuss the details of our approach, and confine ourselves to mentioning an example only. In this example, we assume that $k > 64536$. First, we encode the k source symbols using an extended Hamming Code to produce $\tilde{k}$ symbols. The value of $\tilde{k}$ is roughly $k + \lceil \log_2(k) \rceil$. Next, we use a random left-regular LDPC code with $\tilde{k} + 1000$ message symbols and 580 check symbols in which the message symbols are all of degree 4, and the neighbors of each message symbol are chosen randomly. Using this code, we encode the $\tilde{k}$ symbols into $\tilde{k} + 1000$ input symbols. This completes the description of the pre-code. For the output degree distribution $\Omega(x)$ we choose

$$\begin{aligned}\Omega(x) \quad = \quad & 0.008x + 0.494x^2 + 0.166x^3 + 0.073x^4 + \\ & 0.083x^5 + 0.056x^8 + 0.037x^9 + 0.056x^{19} + \\ & 0.025x^{65} + 0.003x^{66}.\end{aligned}$$

Then we can show that the error probability of the BP decoder on this Raptor Code is at most 1.71×10^{-14}. The proof of this assertion is based on a novel analysis tool for finite length LT Codes [3], combined with an efficient method to analyze right-Poisson LDPC codes on the BEC.

References

[1] M. Luby, "LT-codes," in *Proceedings of the ACM Symposium on Foundations of Computer Science (FOCS)*, 2002.

[2] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielman, "Efficient erasure correcting codes," *IEEE Trans. Inform. Theory*, vol. 47, pp. 569–584, 2001.

[3] R. Karp, M. Luby, and A. Shokrollahi, "Finite length analysis of LT-codes," This Proceedings, 2004.