
Peer reviewed version

[Link to publication record in Explore Bristol Research](http://www.bristol.ac.uk/pure/about/ebr-terms)

[PDF-document](http://www.bristol.ac.uk/pure/about/ebr-terms)
Rateless Distributed Source Code Design

Dino Sejdinovic, Rob Piechocki, Angela Doufexi

Centre for Communications Research, University of Bristol

Kingston-upon-Thames, September 8, 2009.

MobiMedia 2009
Outline

1. Fountain codes: state of the art
2. Rateless coding with side info
3. Fountain coding with multiple source nodes
Outline

1. Fountain codes: state of the art
2. Rateless coding with side info
3. Fountain coding with multiple source nodes
Multicast transmission in a lossy packet network

- Receivers experience different and dynamically changing packet loss rates.
- Wireless erasure networks / mobile environments.
- ARQ/Feedback implosion.
Erasure coding

- Erasure codes (MDS - Reed-Solomon)?
  - low operational complexity (mobile devices: computational resources and battery power)
  - Sparse graph codes coupled with belief propagation (BP) algorithm: LDPC, Turbo, LDGM, IRA...
Erasure coding

- Erasure codes (MDS - Reed-Solomon)?
  - support for a wide range of (and dynamically changing) packet loss rates
    - code rate = ??
Digital fountain

# encoding packets=∞, code rate = 0
Fountain codes are:

- **rateless** - a potentially limitless amount of encoding packets.
- **computationally efficient and scalable** - fast and parallelizable algorithms.
- **nearly optimal** - reliable data reconstruction from any set of encoding packets only slightly greater than the size of the original message.
Digital Fountain’s Raptor FEC has been adopted by:

- **3GPP** Multimedia Broadcast/Multicast
- **DVB-h** IP datacast to handheld devices
- **IETF** Reliable Multicast Transport (RMT)
(Luby 2002), $LT(k, \Omega(x))$ code ensemble: $k$- size of the message, $\Omega(x) = \sum_{d=1}^{k} \Omega_d x^d$ probability distribution on \{1, 2, \ldots, k\} (gen. poly.)

- Sample an output degree $d$ with probability $\Omega_d$.
- Sample $d$ distinct data packets uniformly at random and XOR them.
LT codes achieve capacity

Fact

There exist sequences of LT code ensembles \( LT(k, \Omega^{(k)}(x)) \) which achieve capacity regardless of the erasure probability of the channel (universality) with computational cost of \( \Theta(k \log k) \).

- \( \Omega^{(k)}(x) \) converges pointwise to limiting soliton distribution:
  \[
  \Psi_{\infty}(x) = \sum_{d=2}^{\infty} \frac{x^d}{d(d-1)}
  \]

- Small perturbations suffice at finite lengths: robust soliton.
Soliton distributions
Raptor codes

- (Shokrollahi 2006), $\Omega(x)$ capped at a max. degree $d_{\text{max}}$ as $k \to \infty$ (lowers computational cost to $\mathcal{O}(k)$ but introduces an error floor) - decode fraction $1 - \delta$.
- Error floor is removed by an outer very high rate LDPC code - sufficient redundancy to finish off decoding.
Decoding graph

$\Lambda(x)$

$k$ source packets

$\Omega(x)$

$k(1+\varepsilon)$ encoded packets

$\varepsilon$: code overhead
BP decoding asymptotic analysis

- (Luby, Mitzenmacher, Shokrollahi 1998) AND-OR tree evaluation
- Generalized density evolution techniques (Richardson, Urbanke, MCT, 2008)
- Recipe for fountain code design:
  - formulate a particular version of density evolution - set of recursive equations
  - generate an optimization procedure based on the density evolution equations (typically LP).
Optimisation of $\Omega(x)$

Fix $d_{\text{max}}$ and $\delta$ and minimise $\varepsilon$:

$$\text{LP : } \min \sum_{d}^{d_{\text{max}}} \frac{\omega_d}{d}(\sim 1 + \varepsilon)$$

$$\sum_{d=1}^{d_{\text{max}}} \omega_d (1 - y_i)^{d-1} \geq -\ln y_i, \quad i \in \{1, 2, \ldots, m\}$$

$$\omega_d \geq 0, \quad d \in \{1, 2, \ldots, d_{\text{max}}\}.$$

- $1 = y_1 > y_2 > \cdots > y_m = \delta$ are $m$ equidistant points on $[\delta, 1]$, $\delta$ is the desired error rate, and $d_{\text{max}}$ is the max. degree.
Outline

1. Fountain codes: state of the art
2. Rateless coding with side info
3. Fountain coding with multiple source nodes
The admissible rate region for the pairs of rates \((R_X, R_Y)\) is given by:

\[
\begin{align*}
R_X & \geq H(X|Y) \\
R_Y & \geq H(Y|X) \\
R_X + R_Y & \geq H(X, Y).
\end{align*}
\]
Slepian-Wolf Coding (SWC)

\[ H(Y) \]
\[ H(Y|X) \]
\[ H(X,Y) \]
\[ H(X|Y) \]
\[ H(X) \]
\[ R_x \]
\[ R_y \]

Asymmetric SWC
Symmetric SWC
Admissible rate region
Scalable video multicast

- scalable video over loss-prone wireless networks
- Single channel code for both:
  - video compression (Slepian-Wolf coding)
  - packet loss protection

Rateless Asymmetric SWC

\[ m \geq k H(X \mid Y) \]
"Erasure correlation" SWC

- $Y$ is the output of an erasure channel when $X$ is the input.
- Receivers have a priori knowledge of a number of data packets (transmission from other sources?).
Asymmetric SWC - side information

- Systematic Raptor - *Fresia, Vandendorpe (Globecom 2007).*
- Non-systematic LT (shifted robust soliton) - *Agarwal, Hagedorn, Trachtenberg (ITA Workshop 2008).*
- IR-HARQ with LDPC/Fountain codes - *Sejdinovic, Ponnampalam, Piechocki, Doufexi (IEEE WCNC 2008).*
Fountain codes: state of the art
Rateless coding with side info
Fountain coding with multiple source nodes

Asymptotic code design with side info

$\rho = \frac{\text{number of received symbols}}{k}$
$\zeta = \frac{\text{number of reconstructed symbols}}{k}$

$\Omega(x) = x$, $\Omega(x) = x^2$, $\Omega(x) = x^3$, $\Omega(x) = x^4$, $\Omega(x) = x^5$, $\Omega(x) = x^6$, $\Omega(x) = \Phi_\infty$, $\lambda = 2(x)$

Dino Sejdinovic, Rob Piechocki, Angela Doufexi

Topics in Fountain Coding
Outline

1. Fountain codes: state of the art
2. Rateless coding with side info
3. Fountain coding with multiple source nodes
Multiple source nodes
Rateless Symmetric SWC

\[ X^k \xrightarrow{encoder 1} Z_X^{\infty} \xrightarrow{encoder 2} Z_Y^{\infty} \xrightarrow{decoder} \bar{X} \]

\[ Y^k \]

\[ m \geq \frac{kH(X,Y)}{2} \]
Multiple source nodes

- Obstacles:
  - No cooperation or centralized controller
  - Each source node produces localized encoding packets.

- Questions:
  - How to perform (small) decentralized encoding tasks such that the resulting decoding problem is well-behaved?
  - Can relay help by combining data from multiple sources?
DE: General case

- Packets dispersed across $s$ source nodes.
- Sets of packets available at different nodes are not necessarily disjoint nor of equal size.
- Each source node oblivious of which packets are available at other source nodes.
- *IEEE Commun. Letters 2009* (under review, with Piechocki, Doufexi, Ismail)
Rigorous asymptotic analysis for many data dissemination scenarios.

Generalized DE leads to a simple asymptotic code design yielding both multiterminal source coding and channel coding gains.

Easy modification to include the case of informed collector node, i.e., decoder side information.

Amenable to extension for noisy channels and general belief propagation algorithm (channels like BSC and BIAWGNC).
Example

Source nodes $S_1$ and $S_2$ are trying to multicast $k$ packets. Each source node contains $t > k/2$ packets, but is oblivious of which $t$ packets are available at other node. Source nodes use $LT(t, \Omega(x))$ ensemble and receiver obtains $n/2$ encoding packets from each source node.
Symmetric SWC - DDLT

\[ A_1 \quad \frac{1-p}{2k} \quad 1-p \quad \frac{1}{1+p} \quad \frac{1-n}{2} \quad B_1 \]

\[ A_2 \quad \frac{pk}{2} \quad 1-p \quad \frac{2p}{1+p} \quad \frac{1-n}{2} \quad B_2 \]

\[ A_3 \quad \frac{1-p}{2k} \quad 1-p \quad \frac{2p}{1+p} \quad \frac{1-n}{2} \quad B_3 \]
Optimization of $\Omega(x)$, $p = 1/3$

\[ LP : \quad \min \sum_{d=1}^{d_{\text{max}}} \frac{\omega_d}{d} \]

\[ \frac{1}{1+p} \omega \left( 1 - \frac{1-p}{1+p} y_i - \frac{2p}{1+p} y_i^2 \right) \geq -\ln y_i, \quad i \in \{1, 2, \ldots, m\}, \]

\[ \omega_d \geq 0, \quad d \in \{1, 2, \ldots, d_{\text{max}}\}. \]
Intermediate performance

\[ \rho = \frac{\text{(number of received symbols)}}{k} \]

\[ \zeta = \frac{\text{(number of reconstructed symbols)}}{k} \]

\[ \Omega(x) = x \]

\[ \Omega(x) = x^2 \]

\[ \Psi_\infty(x) \]

Distributions obtained by LP.
Numerical results

- $y_1, y_3 (k=\infty, \Omega^*)$
- $y_2 (k=\infty, \Omega^*)$
- $y_1, y_3 (k=6 \cdot 10^4, \Omega^*)$
- $y_2 (k=6 \cdot 10^4, \Omega^*)$
- $y_1, y_3 (k=1.5 \cdot 10^5, \Omega^*)$
- $y_2 (k=1.5 \cdot 10^5, \Omega^*)$
- $y_1, y_3 (k=6 \cdot 10^4, \Omega_{raptor})$
- $y_2 (k=6 \cdot 10^4, \Omega_{raptor})$
Perturbation of the Limiting soliton for correlated data.

\[
\Omega(x) = -\frac{2(1+p)}{1+3p} \int_0^x \ln \frac{\sqrt{t(p^2 - 1) + 1 - p}}{1 - p} \, dt, \quad x \in [0, 1). \quad (1)
\]

Fact

*When two terminals contain erasure correlated data, fountain coding can still achieve information theoretic limits, provided that the code design is appropriately modified.*
Summary

- Overview of fountain coding - LT, Raptor codes
- DSC with fountain codes
- Generic setting with multiple source nodes
- Symmetric SWC - perturbation of soliton distribution achieves SW limit.