早稲田大学大学院 基幹理工学研究科

博 士 論 文 概 要

論 文 題 目

分散ストレージ符号化の一般化に関する研究

A Study on Generalization of Coding for Distributed Storage System

申 請 者

鎌塚	明
Akira	KAMATSUKA

数学応用数理専攻 情報理論研究

2017 年 10 月

近年，記録ストレージの大容量化や，各種クラウドサービスの発展に伴い，企業および個人が保有する大量のデータを安全かつ効率的に管理する必要性が高まっている．実際，頻繁にアクセスされるデータを保有するストレージについては，故障によるデータ消失のリスクが無視できない頻度で生じ得る．データ消失に対する最も単純な対策としては，保存すべきデータ（元データと呼び $\boldsymbol{m} \in \mathbb{F}_q^B$ で表す．ここで，$\mathbb{F}_q^B$ は位数 q の有限体 $\mathbb{F}_q$ 上の B 次元ベクトル空間を表す．）を複数個（n 個）のストレージに複製することだが，これは元データの n 倍のデータサイズを必要とするため効率が悪い．そこで，元データ $\boldsymbol{m}$ に対して冗長性を付加する写像（符号化と呼ぶ）$\boldsymbol{m} \mapsto (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n) \in (\mathbb{F}_q^\alpha)^n$ を施し，写像されたデータの一部（分散情報と呼ぶ） $\boldsymbol{c}_i, i = 1, \ldots, n$ を各ストレージに保存するシステム（分散ストレージシステム）で元データ $\boldsymbol{m}$ を管理することを考える．ここで，$\alpha \in \mathbb{N}$ を分散情報のサイズと呼ぶ．

分散ストレージシステムが備えるべき主機能は元データの復元機能である．これまで復元機能を持つ符号化として，Reed-Solomon 符号化を始めとする MDS 符号化に関する研究がなされてきた．この符号化により，分散ストレージシステムは

(1) n 個のストレージの内，任意の k 個のストレージが持つ分散情報から元データを復元可能
（$\overset{\text{def.}}{\Longleftrightarrow}$ 任意の $i_1, \ldots, i_k \in \{1, \ldots, n\}$ に対して，ある写像 $(\boldsymbol{c}_{i_1}, \ldots, \boldsymbol{c}_{i_k}) \mapsto \boldsymbol{m}$ が存在）

という機能を実現できる．すなわち，任意の $n-k$ 個のストレージが故障したとしても，残りの k 個のストレージから元データを復元できる．その後，元データの復元機能に加えて，以下のような付加機能を持つ分散ストレージ符号化に関する研究がなされてきた．

(2) 元データの情報漏えい耐性機能
(3) 故障ストレージの効率的な修復機能

(2) を実現する符号化としては，(k, n)-秘密分散法に関する研究がなされてきた．代表的な符号化に Shamir による (k, n)-しきい値法がある．これは，元データと一様乱数をもとに生成した多項式上の n 点を分散情報として各ストレージが保存する方式である．この符号化を用いると，主機能 (1) に加え，

- 任意の $k-1$ 個以下のストレージが持つ分散情報からは元データに関する情報を得られない
（$\overset{\text{def.}}{\Longleftrightarrow}$ 任意の $i_1, \ldots, i_k$ に対して，$H(\boldsymbol{m} \mid \boldsymbol{c}_{i_1}, \ldots, \boldsymbol{c}_{i_{k-1}}) = H(\boldsymbol{m})$），

（ここで $H(\cdot \mid \cdot)$ や $H(\cdot)$ は情報理論における情報エントロピー関数を表す）という機能を分散ストレージシステムに持たせることができる．Shamir は (k, n)-秘密分散法を用いた際の各ストレージが保存すべき分散情報のサイズの限界を示し，(k, n)-しきい値法がこの限界を達成する方式であることを示した．

(3) を実現する符号化としては，近年，Dimakis らによって $[n, k, d]$-再生成符号化が提案されている．従来，故障ストレージの修復は，復元した元データに再度符号化を施すことによってなされてきた（これを自明な修復法と呼ぶ）．分散情報のサイズが α であるため，自明な修復法には $k\alpha$ だけの通信量が必要になる．$[n, k, d]$-再生成符号化された分散ストレージシステムにおいては，j 番

目の故障ストレージの修復の際にはまず，$d\,(\leq n-1)$ 個の修復用ストレージが選ばれる．選ばれた各ストレージは，各々の分散情報 $\boldsymbol{c}_i$ から修復用データ $p_{i\to j}\in\mathbb{F}_q^{\beta}$ を生成（$\boldsymbol{c}_i\mapsto p_{i\to j}$）し，故障ストレージに送信する．故障ストレージは d 個の修復用データを用いて分散情報を再生成することにより修復を行う．修復用データのサイズは $\beta\in\mathbb{N}$ であるので，このときの通信量（修復バンドワイズと呼ぶ）は $d\beta(\leq k\alpha)$ となり，自明な修復法よりも効率的に修復ができる．Dimakis らは，$[n,k,d]$-再生成符号における分散情報のサイズ α と修復バンドワイズ $d\beta$ の間のトレードオフ不等式を示した：

$$\sum_{i=0}^{k-1}\min\{\alpha,(d-i)\beta\}\geq B,$$

ここで，B は元データのサイズを表す．このトレードオフ不等式において，分散情報のサイズ α を最小にしたもとで修復バンドワイズ $d\beta$ を最小にする $[n,k,d]$-再生成符号化を MSR (minimum-storage regenerating) 符号化と呼ぶ．一方，修復バンドワイズを最小にしたもとでストレージを最小にする $[n,k,d]$-再生成符号化を MBR (minimum-bandwidth regenerating) 符号化と呼ぶ．具体的な MSR/MBR 符号化の構成法としては，Rashmi らによる Product Matrix 法や，Shah らによる Repair by Transfer 法等が提案されている．

上記の分散ストレージシステムの機能 (1)(2)(3) はいずれも，ストレージ数が一定のしきい値（k や d）以上あるいは以下になったときに発揮されるしきい値型の分散ストレージシステムである．しかしながら，実際の分散ストレージシステムの構築および運用においては，すべてのストレージが全く同じ能力（ストレージ容量，耐久性，計算能力等）を持っているわけではないため，ストレージ毎の役割を考慮した分散ストレージを設計することが必要になる．すなわち，(1)(2)(3) の条件を一般化した機能をもつ分散ストレージステムの構築が必要である．

(1) および (2) の条件の一般化に関する研究としては，伊東らによって Γ-秘密分散法が提案されている．ここで，Γ は (1) および (2) に関して一般化された条件を表す．具体的には，n 個のストレージのインデックスを表す集合を $\{1,\ldots,n\}$ とするとき，元データを復元可能なストレージのインデックスの集合族 $\mathcal{A}$ と，元データに関する情報を一切得られないストレージのインデックスの集合族 $\mathcal{B}$ の組 $\Gamma=(\mathcal{A},\mathcal{B})$ として定義される．Γ-秘密分散法の構成法としては，各ストレージに対して，しきい値法によって生成される分散情報を複数個割り当てる方式（複数割当法（ふくすうわりあてほう））が提案されている．ここで，与えられた条件 Γ を満たすためには，ストレージ集合 $\boldsymbol{A}\in\mathcal{A}$ に対しては，ある (t,m)-しきい値法で復元するのに十分な個数（t 個以上）の分散情報を割り当て，$\boldsymbol{B}\in\mathcal{B}$ に対しては，(t,m)-しきい値法で元データに関する情報が得られなくなるような個数（$t-1$ 個以下）の分散情報を割り当てればよい．その後，各ストレージが保存する分散情報サイズの平均 ρ を最小化する構成法が岩本らによって提案されている．この構成法は $\mathcal{A},\mathcal{B}$ に対する割当の仕方を制約として，ρ を最小化する整数計画問題を繰り返し解くことで最適なパラメータ (t,m) を探索している．

本研究では，(1) および (3) に関する条件を一般化した Ω-再生成符号およびその構成法を提案する．この一般化の動機づけとしては例えば，「耐久性の高いストレージを，故障ストレージの修復に多く参加させたい場合」が挙げられる．本研究では (1) および (3) に関する一般化条件 Ω を

ストレージのインデックスの集合族の組 $\Omega = \left(\mathcal{A}, (\mathcal{B}_j)_{j=1}^n\right)$ として定義し，Ω が従来の $[n,k,d]$-再生成符号における条件 (1) および (3) を含むことを示す．ここで，$\mathcal{A}$ は元データを復元可能なストレージのインデックスの集合族を表し，$\mathcal{B}_j$ は故障ストレージ j を修復可能なストレージのインデックスの集合族を表す．

Ω-再生成符号においては，各ストレージ i が保存する分散情報 $\boldsymbol{c}_i \in \mathbb{F}_q^{\alpha_i}$ のサイズ $\alpha_i \in \mathbb{N}$ や，故障ストレージ j へ送信する修復用データ $p_{i\to j} \in \mathbb{F}_q^{\beta_{i\to j}}$ のサイズ $\beta_{i\to j} \in \mathbb{N}$ が，ストレージごとに異なる．そこで本研究では，Ω-再生成符号の評価基準として，各ストレージが保存する分散情報のサイズの平均 ρ_S および修復バンドワイズの平均 ρ_R を提案する．

Ω-再生成符号の構成法としては，従来の再生成符号の分散情報を用いた複数割当法を提案する．このとき，与えられた条件 Ω を満たすための条件として，$\boldsymbol{A} \in \mathcal{A}$ に対しては，ある $[\ell,t,r]$ 再生成符号で元データを復元するのに十分な個数（t 個以上）の分散情報を割り当て，$\boldsymbol{B} \in \mathcal{B}_j$ に対しては，$[\ell,t,r]$-再生成符号において故障ストレージ j を修復するのに十分な個数（r 個以上）の分散情報を割り当てればよいことを示す．

本研究ではさらに，複数割当法による符号クラスの中で「ρ_S を最小にしたもとで ρ_R を最小にする符号 (Ω-MSR-map 符号)」および「ρ_R を最小にしたもとで ρ_S を最小にする符号 (Ω-MBR-map 符号)」を定義し，その構成アルゴリズムを導出する．この構成アルゴリズムではまず，$\mathcal{A}, (\mathcal{B}_j)_{j=1}^n$ に対する割当の仕方を制約とした，ρ_S あるいは ρ_R のいずれか一方を最小化する整数計画問題を繰り返し解き，最小値を与えるパラメータ $[\ell,t,r]$ を探索する．次に，探索したパラメータの中でもう一方を最小化する $[\ell,t,r]$ を求める．ここで，Ω-MSR/MBR-map 符号を構成する際には，Γ-秘密分散法の場合と異なり，$[\ell,t,r]$-再生成符号におけるパラメータ (α,β) をも最適化する必要がある．本提案においては，それぞれ MSR/MBR 符号のパラメータを用いれば良いことを示す．

Ω-再生成符号は Γ-秘密分散法と異なり (3) の修復条件を一般化しているため，故障ストレージの修復法に関して工夫の余地がある．そこで本研究ではさらに，複数割当法を用いた場合の，通信量の意味でより効率的な修復法を提案し，その効率性について解析を行う．また，その修復法を用いた場合の Ω-MSR/MBR-map 符号の構成法についても考え，Ω-MSR-map 符号については上述と同様の構成法が導出できることを示す．一方，Ω-MBR-map 符号については，割り当てる再生成符号における最適なパラメータ (α,β) が決定できないため，準最適な構成法を提案する．

本論文の構成は以下の通りである．まず第 1 章で本研究の背景および目的について述べる．第 2 章では，準備として $[n,k,d]$-再生成符号および (k,n)-秘密分散法について概観する．第 3 章では，復元および再生成に関する条件を一般化した Ω-再生成符号とその評価基準を提案する．構成法としては複数割当法を提案し，複数割当法を用いた符号クラスの中での最適な符号として，Ω-MSR/MBR-map 符号の構成アルゴリズムを導出する．第 4 章では 複数割当法を用いた場合の故障ストレージの修復法に関して，通信量の意味でより効率的な修復法を提案し，修復にかかる通信量について解析を行う．そのうえで，効率的な修復法を用いた場合の Ω-MSR/MBR-map 符号の構成アルゴリズムについて考察する．第 5 章では，提案した各アルゴリズムに関して，具体的な数値例を構成し，効率性について考察する．最後に，第 6 章で本論文の結論と今後の展望を述べる．

早稲田大学　博士（工学）　学位申請　研究業績書

氏 名　　鎌塚 明　　印

（2017　年　10 月　現在）

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
1.論文〇	復元および再生成の条件を一般化した再生成符号とその構成法，電子情報学会論文誌 A, 2017，Vol.J100-A,No.11,pp.-,Nov. 2017.（掲載決定） 鎌塚 明，東 優太，吉田 隆弘，松嶋 敏泰
2.講演	“Regenerating codes with generalized condi- tions of reconstruction and regeneration,” in 2016 International Symposium on Information Theory and Its Applications (ISITA), pp. 41--45, Oct 2016, A. Kamatsuka, Y. Azuma, T. Yoshida, and T. Matsushima
3.講演	“A maximum likelihood decoding algorithm of gabidulin codes in deterministic network coding,” in 2016 International Symposium on Information Theory and Its Applications (ISITA), pp. 666--670, Oct 2016, K. Kazama, A. Kamatsuka, and T. Matsushima
4.講演	"Parallel Concatenation of Polar Codes and Iterative Decoding,” Proceedings of the 2014 International Symposium on Information Theory and its Applications, p.347, Oct 2014, Akira KAMATSUKA, Shunsuke HORII, Toshiyasu MATSUSHIMA
5.講演	一般化された再生成符号に対する効率的な複数割当法による構成法，第 39 回情報理論とその応用シンポジウム（SITA2016），富山県，2016 年 12 月， 鎌塚 明，吉田 隆弘，松嶋敏泰
6.講演	潜在変数を仮定した非線形回帰モデルおけるベイズ基準のもと最適な予測，第 39 回情報理論とその応用シンポジウム（SITA2016），富山県，2016 年 12 月， 鎌塚 明，吉田 隆弘，松嶋敏泰
7.講演	シンボルペア通信路における符号のリスト復号に関する一考察，第 39 回情報理論とその応用シンポジウム（SITA2016），富山県，2016 年 12 月， 風間皐希，鎌塚 明，松嶋敏泰
8.講演	Array-Error モデルにおける軟判定復号に関する一考察,電子情報通信学会技術研究報告,vol.115, no.394, IT2015-49, pp.7--12,電子情報通信学会情報理論研究会（IT），2016 年 1 月 18 日～19 日，大阪府，　風間皐希，鎌塚明，松嶋敏泰
9.講演	条件を一般化した再生成符号とその複数割当法による構成法，第 5 回誤り訂正符号のワークショップ(ECCW2016)，佐賀県，2016 年 9 月 鎌塚 明，吉田隆弘，松嶋敏泰
10.講演	Polar 符号の探索アルゴリズムを用いた復号法に関する一考察，第 38 回情報理論とその応用シンポジウム（SITA2015），岡山県，2015 年 11 月 鎌塚 明，松嶋敏泰

早稲田大学　博士（工学）　学位申請　研究業績書

種 類 別	題名、　発表・発行掲載誌名、　発表・発行年月、　連名者（申請者含む）
11.講演	復元および再生成の条件を一般化した再生成符号に関する一考察, 第 38 回情報理論とその応用シンポジウム（SITA2015）, 岡山県, 2015 年 11 月 東 優太, 鎌塚 明, 吉田 隆弘, 松嶋敏泰
12.講演	Polar 符号の探索アルゴリズムを用いた復号について, 電子情報通信学会情報理論研究会（IT）, 石川県, 2015 年 9 月, 鎌塚 明, 松嶋敏泰
13.講演	消失中継通信路上での Decode - and - Forward 型通信におけるパンクチャされた空間結合 LDPC 符号のユニバーサル性, 第 37 回情報理論とその応用シンポジウム（SITA2014）, 岡山県, 2014 年 12 月 中原悠太, 齋藤翔太, 鎌塚 明, 松嶋敏泰
14.講演	非線形コンバイナ型乱数生成器に対する Sum - Product Algorithm を用いる攻撃に関する一考察,電子情報通信学会技術研究報告,vol.114, no.306, IBISML2014-83, pp.357-364, 電子情報通信学会情報論的学習理論と機械学習研究会（IBISML）, 2014 年 11 月 17 日～19 日, 愛知県, 久保航汰, 齋藤翔太, 鎌塚明, 松嶋敏泰
15.講演	有限バッファ Hybrid SR-ARQ における最適制御方式に関する一考察,電子情報通信学会情報理論研究会（IT）, 兵庫県, 2014 年 7 月, 影山優太, 鎌塚 明, 前田康成, 松嶋敏泰
16.講演	Polar 符号を用いた並列連接符号化に関する一考察, 第 36 回情報理論とその応用シンポジウム（SITA2013）, 岡山県, 2013 年 11 月, 鎌塚 明, 堀井俊佑, 松嶋敏泰