早稲田大学大学院　基幹理工学研究科

博 士 論 文 審 査 報 告 書

論　文　題　目

計算時間，信頼性，情報秘匿性を考慮した分散符号化計算方式に関する研究

A Study on Distributed Coded Computation Methods Considering Computation Times, Error-Correcting Capabilities and Information Securities

申　請　者

風間　皐希

Koki KAZAMA

数学応用数理専攻　情報理論研究

2022年2月

大規模データを扱う機会が増えた現代社会において，高速なデータ処理方式の一つである分散計算方式の必要性がいっそう高まっている．これは，主要な計算機（マスター）が複数の計算機（ワーカー）にデータの一部を分散し，並列計算を行わせる方式である．しかし，これには以下の (a) の利点と (b) や (c) の欠点がある．

(a) システムの総計算時間が短くなる．

(b) ワーカーの故障等により，一部のワーカーが誤った計算値をマスターに返す，または計算結果を返さない可能性が，ワーカーの数の増加とともに高まる．（本論文では，ワーカーが誤った計算値を返すことを誤りが発生した，計算結果を返さないことを消失が発生したと表現している．）

(c) データに個人情報などユーザの秘密情報が含まれる場合，データをワーカーに分散することで秘密情報がワーカーに漏洩する可能性が高まる．

本論文では，分散計算方式に誤り訂正符号を導入することで，利点 (a) を有しつつ，欠点 (b)(c) を解消する方式である分散符号化計算方式を論じている．本論文の内容は，主に第 4 章で論じられる (a)(b) を考慮した方式の研究と，第 5 章で論じられる (a)(b)(c) 全てを考慮した方式の研究の 2 つで構成されている．前者では，従来から (a)(b) を考慮した分散符号化計算方式が提案されているが，本論文では (b) の誤り消失訂正の範囲を拡張した方式を提案している．一方，後者では，(a)(b)(c) 全てを同時に考慮した新たな問題設定における方式を提案している．従来研究として，(b)(c) を考慮した秘匿分散符号化計算方式が研究されているが，(a) については考慮されていなかった点で本論文とは異なっている．

本論の構成は，6 章から成り立っている．第 1 章では，本論文の位置づけを論じている．第 2 章では，本論文で用いる有限体，確率論，情報理論の数理的基礎をまとめ，基礎理論となる誤り訂正符号と情報セキュリティの秘密分散法について概説している．

第 3 章では従来研究の分散符号化計算方式と秘匿分散符号化計算方式の概要を述べ，本論文で取り扱う問題とプロトコルについてまとめている．従来研究では，マスターとワーカーについて抽象化された計算機を仮定し，実機のハードウエアの詳細な構成等までには踏み込まず，(a)(b)(c) の評価を理論的に論じている．従来の主要な問題設定では，位数 q の有限体 $\mathbb{F}_q$ 上の比較的巨大な $k_{\mathrm{A}} \times l$ 行列 $\boldsymbol{A}$ と $l \times k_{\mathrm{B}}$ 行列 $\boldsymbol{B}$ の積行列 $\boldsymbol{AB}$ の計算方式が研究されている．基本的プロトコルとしては，行列 $\boldsymbol{A}$ は固定値，行列 $\boldsymbol{B}$ は可変値として，毎回 $\boldsymbol{B}$ は異なった値が入力され，その都度 $\boldsymbol{AB}$ の計算値を出力する問題を扱っている事が多い．本論文でも基本的には上記の問題とプロトコルについて論じている．

従来研究の共通した分散符号化計算の手順は，まず前処理として固定値の行列 $\boldsymbol{A}$ を符号化により変換した行列 $\tilde{\boldsymbol{A}}$ を求め，それをある部分行列に分割し，各ワーカーに分散して保存させておく．毎回入力される行列 $\boldsymbol{B}$ はマスターから全てのワーカーに送信され，ワーカーはそれぞれ保存していた $\tilde{\boldsymbol{A}}$ の部分行列と行列 $\boldsymbol{B}$ を用いある計算を行い，その計算結果をマスターに送信する．各ワーカーから計算結果が送信されない，または誤った計算結果が送信される事象が生じるが，マスターはそれらの情報を集約し復号化による変換により，正しい計算結果である積行列 $\boldsymbol{AB}$ を求めることを試みる．本研究でもこの基本的手順は踏襲しながらも，訂正範囲の拡張や新たな機能を追加している．

第 4 章では，まず第 3 章で述べた従来の分散符号化計算方式を再定式化することで，誤り訂正符号の原理が機能する部分を明確化している．この再定式化により，従来の方式は，ワーカーの分散計算結果を集約した行列が，求めたい積行列 $\boldsymbol{AB}$ を列ごとに誤り訂正符号化した行列となるよう，処理が行われていることを明らかにしている．この章では，ワーカーの分散計算結果を集約した行列が，積行列 $\boldsymbol{AB}$ 自体を一括に誤り訂正符号化した行列となるように一般化した分散符号化方式を提案している．具体例としては $\boldsymbol{AB}$ を一括に Gabidulin 符号化する方式の構成法を提案している．この新方式は，従来法では訂正不可能であった行列のランク誤りに対しても耐性をもつ分散符号化方式となっている．

Gabidulin 符号を用いる方式の手順を概説する．正整数 $n, n_{\mathrm{A}}, n_{\mathrm{B}}$ は，$n > k_{\mathrm{A}}$，$m \coloneqq \max\{n, k_{\mathrm{B}}\}$，$n_{\mathrm{A}}|n$，$n_{\mathrm{B}}|m$ を満たす．$n_{\mathrm{A}}n_{\mathrm{B}}$ 台ある全ワーカーは n_{A} 個のグループに均等に分かれている．

(1) 行列 $\boldsymbol{A}$ が入力されると，マスターは，前処理として，$\boldsymbol{A}$ を符号化した行列 $\tilde{\boldsymbol{A}}$ を行方向に n_{A} 分割した i 番目の部分行列 $\tilde{\boldsymbol{A}}_{i\cdot}$ を各グループ $i \in [n] \coloneqq \{1, \ldots, n\}$ 内の全ワーカーに保存させる．

(2) 行列 $\boldsymbol{B}$ がマスターから全ワーカーに送信されると，各グループ i は，$\tilde{\boldsymbol{A}}_{i\cdot}$ と $\boldsymbol{B}$ からグループ内で並列計算を行い $\tilde{\boldsymbol{A}}_{i\cdot}(\boldsymbol{B}, \boldsymbol{0}_{l\times(m-k_{\mathrm{B}})})\boldsymbol{v} \in \mathbb{F}_{q^m}^{n/n_{\mathrm{A}}}$ を $\mathbb{F}_q$ 上行列表示したもの $\boldsymbol{f}^{n/n_{\mathrm{A}}}(\tilde{\boldsymbol{A}}_{i\cdot}(\boldsymbol{B}, \boldsymbol{0}_{l\times(m-k_{\mathrm{B}})})\boldsymbol{v}) \in \mathbb{F}_q^{(n/n_{\mathrm{A}})\times m}$ を得る．ただし，$\boldsymbol{v} \in \mathbb{F}_{q^m}^m$ は $\mathbb{F}_q$ 上線形空間 $\mathbb{F}_{q^m}$ の基底を並べたベクトルである．

(3) マスターは，全ワーカーからの受信結果から得た $\boldsymbol{f}^n(\tilde{\boldsymbol{A}}(\boldsymbol{B}, \boldsymbol{0}_{l\times(m-k_{\mathrm{B}})})\boldsymbol{v}) + \boldsymbol{E} \in \mathbb{F}_q^{n\times m}$ を復号し，$\boldsymbol{AB}$ の推定計算結果を得る．ただし，$\boldsymbol{f}^n(\tilde{\boldsymbol{A}}(\boldsymbol{B}, \boldsymbol{0}_{l\times(m-k_{\mathrm{B}})})\boldsymbol{v})$ は $\tilde{\boldsymbol{A}}(\boldsymbol{B}, \boldsymbol{0}_{l\times(m-k_{\mathrm{B}})})\boldsymbol{v} \in \mathbb{F}_{q^m}^n$ を $\mathbb{F}_q$ 上行列表示したものであり，$\boldsymbol{E}$ は計算中の誤りを表す行列である．

この具体的提案方式について (a)(b) に関する性能評価を以下のように論じている．

定理 1((b) に関する評価（概要）) $\mathrm{rank}\boldsymbol{E} \leq t$ ならば，本方式で $\boldsymbol{AB}$ が正しく計算される．

以上より従来研究では訂正不可能な誤りも提案方式で訂正可能となっていることが示されている．

(a) の計算時間の評価について，本論文では，各ワーカーの計算処理とマスターの復号処理における $\mathbb{F}_q$ 上四則演算回数を計算時間と定義している．例えば，マスター単独で $\boldsymbol{AB}$ を計算する方式 (単独方式) では，$k_{\mathrm{A}}k_{\mathrm{B}}(2l-1)$ 回となり，これより提案方式の計算時間が短いことが望ましい性質となる．

定理 2((a) に関する評価（概要）) 仮定として，q は 2 のべき乗としたもとで，$\log_2 q$ と m は互いに素，かつ $2m+1$ は素数，かつ乗法群 $(\mathbb{Z}/(2m+1)\mathbb{Z})^\star$ は 2 と -1 から生成されるとする．この仮定のもとで，本提案方式の計算時間の上界 T_1 は $T_1 = (2k_{\mathrm{B}}l-1)(mn/n_{\mathrm{A}}n_{\mathrm{B}}) + D$ で与えられる．ただし，$t = \lfloor (n-k_{\mathrm{A}})/2 \rfloor$，$d = n - k_{\mathrm{A}} + 1$，$D = (2mnt + m^2 + m - 1 + \frac{2}{3}(m(m-1)(2m-1) - t(t-1)(2t-1)) - (t-2)(m-t) - (t-1)(m^2 - m - t^2 + t)) + (dn + d^2 - t^2 + 2dt + mt - n - 4d - t - 1)m + (dn + 3d^2 + 3t^2 - 4dt + mt - n - 9d + 9t + 5)(2m^2 - 1) + 2t(2m^2 - 1)(2\lfloor \log_2(m \log_2 q - 1) \rfloor)$ と定義する．

この定理より，単独方式と比較し提案方式が有効となる以下の条件が求まる．$l > \frac{1}{2k_{\mathrm{B}}(k_{\mathrm{A}} - (mn/n_{\mathrm{B}}n_{\mathrm{A}}))}(k_{\mathrm{A}}k_{\mathrm{B}} - (mn/n_{\mathrm{A}}n_{\mathrm{B}}) + D)$.

第 5 章では，本論文の新たな方式として，(a)(b)(c) を同時に考慮した分散符号化計算方式を提案し，その性能評価をしている．(c) の秘匿性の条件は，行列 $\boldsymbol{A}$ の値を，一部のワーカーは結託し自身の有する情報を持ち寄り $\boldsymbol{A}$ の値を得ようとしても得られないとしている．なお，行列 $\boldsymbol{B}$ の値は公開してもよい．第 5 章で論じられている基本提案方式の概要は以下のような方式となっている．

正整数 $n_{\mathrm{A}}, h_{\mathrm{A}}, n_{\mathrm{B}}$ を，$q > n_{\mathrm{A}} \geq h_{\mathrm{A}}, n_{\mathrm{B}}|k_{\mathrm{B}}$ を満たすようにとる．$n_{\mathrm{A}}n_{\mathrm{B}}$ 台のワーカーは n_{A} 個のグループに均等に分かれている．$\boldsymbol{A} \in \mathbb{F}_q^{k_{\mathrm{A}}\times l}$ は一様分布に従う確率変数行列であるとする．

(1) 行列 $\boldsymbol{A}$ が入力されると，マスターは，前処理として，行列 $\boldsymbol{A}$ と，$\boldsymbol{A}$ とは独立に一様分布に従う確率変数行列 $\boldsymbol{R} \in \mathbb{F}_q^{k_{\mathrm{A}}(h_{\mathrm{A}}-1)\times l}$ を並べた行列を符号化，分割した部分行列 $\tilde{\boldsymbol{A}}_{i\cdot} \in \mathbb{F}_q^{k_{\mathrm{A}}\times l}$ を得る．グループ i の全ワーカーはこれを保存する．

(2) 行列 $\boldsymbol{B}$ がマスターから全ワーカーに送信されると，グループ i は $\tilde{\boldsymbol{A}}_{i\cdot}\boldsymbol{B}$ をグループ内で並列計算してマスターに送信する．ここで，マスターは一部のグループからは受信しなかったとする．

(3) マスターは，各ワーカーから受信した計算結果を復号して，$\boldsymbol{AB}$ の推定計算結果を得る．

この基本提案方式における (a)(b)(c) に関する性能評価を以下のように論じている．

定理 3((c) に関する評価（概要）) $h_{\mathrm{A}} - 1$ 個のグループ $i_1, \ldots, i_{h_{\mathrm{A}}-1} \in [n_{\mathrm{A}}]$ に属する全てのワーカーが結託しても $\boldsymbol{A}$ の情報が 1bit も漏洩しない，すなわち，$\mathrm{I}(\boldsymbol{A}\,;\, \tilde{\boldsymbol{A}}_{i_1\cdot}, \ldots, \tilde{\boldsymbol{A}}_{i_{h_{\mathrm{A}}-1}\cdot}) = 0$ が成立する．

定理 4((b) に関する評価（概要）) h_{A} 個のグループ $i_1, \ldots, i_{h_{\mathrm{A}}} \in [n_{\mathrm{A}}]$ の計算結果 $\tilde{\boldsymbol{A}}_{i_1 \cdot}\boldsymbol{B}, \ldots, \tilde{\boldsymbol{A}}_{i_{h_{\mathrm{A}}} \cdot}\boldsymbol{B}$ が得られれば $\boldsymbol{AB}$ の値を正しく計算できる，すなわち，$\mathrm{H}(\boldsymbol{AB} \mid \tilde{\boldsymbol{A}}_{i_1 \cdot}\boldsymbol{B}, \ldots, \tilde{\boldsymbol{A}}_{i_{h_{\mathrm{A}}} \cdot}\boldsymbol{B}) = 0$ が成立する．

定理 5((a) に関する評価（概要）) 基本提案方式の計算時間 T_2 は以下となる．
$T_2 = \frac{k_{\mathrm{A}} k_{\mathrm{B}} (2l-1)}{n_{\mathrm{B}}} + k_{\mathrm{A}} k_{\mathrm{B}} (2h_{\mathrm{A}} - 1) + 2h_{\mathrm{A}}^2 - 1.$

この定理より，単独方式と比較し提案方式が有効となる以下の条件が求まる．$(k_{\mathrm{B}} \geq) n_{\mathrm{B}} > (2l - 1)/(2l - 2h_{\mathrm{A}} - (2h_{\mathrm{A}}^2 - 1)/(k_{\mathrm{A}} k_{\mathrm{B}}))$ かつ $1 \leq h_{\mathrm{A}} < (-k_{\mathrm{A}} k_{\mathrm{B}} + \sqrt{k_{\mathrm{A}}^2 k_{\mathrm{B}}^2 + 4k_{\mathrm{A}} k_{\mathrm{B}} l + 2})/2.$

さらに，第 5 章では，グループ内で発生した消失も訂正可能な方式など，情報秘匿性や信頼性を一般化した方式の構成法も提案している．それらについても，(a)(b)(c) の性能評価を行うことで，これらの方式が 3 つの評価基準に対して優れた特性を併せ持つ方式であることを明らかにしている．

最後に第 6 章では，本論文の結論と今後の課題を説明している．

以上を総括する．本論文は，(a) 計算時間，(b) 信頼性，(c) 情報秘匿性という評価基準の下で良い性質を有する分散符号化計算方式の構成法に関する理論研究である．

前半では，(a)(b) の基準で良い性質を有する新たなグループ型分散符号化計算方式の構成法を提案している．提案方式では，求めたい積行列 $\boldsymbol{AB}$ に $\mathbb{F}_{q^m}$ 上 (n, k_A) 符号化を行うのと同等な処理を行うことにより，(b) については 従来研究では不可能な範囲の誤り訂正を可能にすると同時に，(a) については $\mathbb{F}_q$ 上の計算に分解することや復号法の工夫で高速な並列処理を実現し計算時間も短縮している．

後半では，(a)(b)(c) の基準で良い性質を有するグループ型分散秘匿符号化計算方式の構成法を提案している．(b)(c) については，秘密分散法や最大距離分離符号の性質を用い信頼性と情報秘匿性を同時に満たすもとで，(a) については計算の並列化を行うことで計算時間の短縮化も実現している．

以上，これらの成果は分散符号化計算方式の研究分野において，新たな知見をあたえる意義ある成果と言え，本論を博士（工学）の学位論文として価値のあるものと認める．

2022 年 1 月

審査員

（主査）　早稲田大学教授　博士（工学）（早稲田大学）　松嶋　敏泰

早稲田大学教授　工学博士　（早稲田大学）　大石　進一

早稲田大学教授　博士（工学）（早稲田大学）　柏木　雅英

早稲田大学名誉教授　工学博士　（大阪大学）　平澤　茂一